UNIVERSIDADE DE LISBOA

FACULDADE DE LETRAS

VOICE AND SPEECH PERCEPTION IN AUTISM:

A SYSTEMATIC REVIEW

Svetlana Postarnak

Tese orientada pela Prof.ª Doutora Ana Patrícia Teixeira Pinheiro e co-orientada pela Prof.ª Doutora Diana Maria Pinto Prata, especialmente elaborada para a obtenção do grau de Mestre em Ciência Cognitiva

2017

"[…] the task of cognitive science is to characterise the brain, not at the level of nerve cells, nor at the level of conscious mental states, but rather at the level of its functioning as an information processing system."

(John Searle, *Minds, Brains and Science*, pp.43)

ABSTRACT

Autism spectrum disorders (ASD) are characterized by persistent impairments in social communication and interaction, restricted and repetitive behavior. In the original description of autism by Kanner (1943) the presence of emotional impairments was already emphasized (self-absorbed, emotionally cold, distanced, and retracted). However, little research has been conducted focusing on auditory perception of vocal emotional cues, being the audio-visual comprehension most commonly explored instead. Similarly to faces, voices play an important role in social interaction contexts in which individuals with ASD show impairments. The aim of the current systematic review was to integrate evidence from behavioral and neurobiological studies for a more comprehensive understanding of voice processing abnormalities in ASD. Among different types of information that the human voice may provide, we hypothesize particular deficits with vocal affect information processing by individuals with ASD. The relationship between vocal stimuli impairments and disrupted Theory of Mind in Autism is discussed. Moreover, because ASD are characterized by deficits in social reciprocity, further discussion of the abnormal oxytocin system in individuals with ASD is performed as a possible biological marker for abnormal vocal affect information processing and social interaction skills in ASD population.

Key-words: Vocal, Speech, Auditory, Emotion, Oxytocin, Theory of Mind

TABLE OF CONTENTS

1. INTRODUCTION

In our daily lives, we are surrounded by thousands of sounds, but while we automatically pay attention to some of them, others are disregarded. Among different types of information that an auditory signal may provide, vocal sounds seem to be the most relevant to human beings (Belin, Zatorre, & Ahad, 2002; Blasi et al., 2015) Human vocalizations and speech are socially orienting signals (Abrams et al., 2013) that are processed effortlessly during social interactions and communication and to which human's attentional system seems to be always tuned to.

Humans are born to become social beings (Tsao, 2008), who start to develop social competence and language skills from early infancy. When development does not progress within typical stages, these cognitive abilities may suffer some impairment, with important consequences for interpersonal relationships establishment. Atypical social interaction and communication reciprocity is one of the main characteristics of autism spectrum disorders (ASD), in which symptoms manifest along the first years of life (American Psychiatric Association, 2013). The absence of a neurobiological marker prevents the detection of ASD before birth, and thus hampers earlier intervention strategies. Furthermore, earlier intervention strategies are difficult to be adapted as a spectrum disorders rely on the different levels of severity, which are difficult to disentangle. Recently, one more diagnostic criterion was added as characteristic of this neurodevelopmental disorder: the hyper or hyposensitivity to sensory stimuli (American Psychiatric Association, 2013). Indeed, reduced auditory preference towards socially relevant vocal stimuli has been progressively reported in ASD individuals (Bidet-Caulet et al., 2017; Gervais et al., 2004). However, it remains to be clarified why individuals with ASD show early absence of auditory orienting towards human vocalizations and speech stimuli, which serve as reliable cues for socially relevant events.

The current systematic review represents a quantitative analysis of the evidence coming from different scientific fields, including psychology, linguistics and neuroscience, examining how different types of information conveyed by the voice are processed by individuals with ASD compared to typically developing individuals. Moreover, a comparative analysis is produced in order to understand how individuals with ASD differ from typically developing individuals in the processing of vocal information, and

which type of vocal information causes main perceptual difficulties. A collection of scientific research was performed using the PRISMA guidelines: a qualified scientific tool with a set of rigorous criteria, which allows a careful selection of relevant studies, crucial for a systematic review (Moher, Liberati, Tetzlaff, Altman, & Group, 2009). Moreover, the conceptualized model in auditory neuroscience of voice perception proposed by Belin and collaborators (2004) was followed in order to organize the behavioral, neurobiological and electrophysiological evidence about vocal information processing in individuals with autism. According to this model, there are partially dissociated cerebral pathways for the processing of the three types of information that humans can decode from a voice: identity, speech and affect (Belin, Fecteau, & Bédard, 2004). Therefore, the separation of the cerebral pathways in the proposed model is useful for the analysis and understanding of which type of vocal information may be more impaired in individuals with ASD.

A brief analysis of the scientific outcomes of intranasal oxytocin in ASD is performed, as oxytocin dysregulation is one of potential explanations for abnormal vocal affect information processing. Indeed, the observed low endogenous levels of this neuropeptide in individuals with ASD (Kumar, von Kriegstein, Friston, & Griffiths, 2015; Yatawara, Einfeld, Hickie, Davenport, & Guastella, 2016) suggests that oxytocin dysregulation may serve as an important biological marker for autism. The release of oxytocin occurs when typically developing infants are listening to their mother's vocalizations (Gordon et al., 2016), which is important in the establishment of parent-child bounding (Quattrocki & Friston, 2014). In addition, the reduced auditory preference towards human sounds and mother's voice in particular in individuals with ASD may reflect the impaired recognition of different types of vocal information, including the vocal affect information. The recognition of emotional value is important to accurately attribute and understand mental states; a cognitive ability known as the Theory of Mind, which seems to be disrupted in individuals with ASD (Baron-Cohen, 2001). A brief discussion of this explanatory hypothesis of ASD is also produced in order to understand the possible relationship between the vocal information processing and the impaired comprehension of mental states.

1.1 What's in a Voice?

Before newborn infants acquire their native language and build a mental lexicon, they are exposed to a heterogeneous auditory environment, surrounded by human voices as well as non-social auditory stimuli (Grossmann, et al., 2010). However, in spite of this "auditory chaos", infants show a sophisticated voice perception system with preference for vocal stimuli as compared to non-vocal environmental sounds (Blasi et al., 2015), as well as the capacity to discriminate different voices and recognize their parent's voices (Belin & Grosbras, 2010), with a marked listening preference for their mother's voice as compared to other female voices (DeCasper & Fifer, 1980). Thus, the infant's auditory preference for human vocal sounds since birth shows the special salience of this type of sounds among other sounds of the auditory landscape (Belin et al., 2004).

Neuroimaging evidence shows that the adult human brain contains specialized regions that are not only sensitive, but also strongly selective to human voices (Belin et al., 2000). These voice regions, known as temporal voice areas (TVAs), are mostly located along the middle and anterior parts of the superior temporal sulcus (see Appendix) predominantly in the right hemisphere (Belin & Grosbras, 2010). In typically developing infants, the cerebral specialization of TVAs emerges between 4 and 7 months of age (Grossmann et al., 2010), which converges with the myelination period of the primary auditory cortex - around 6 months of age, suggesting a rapid structural change that supports the specialization for voice processing during the first year of life (Belin & Grosbras, 2010). The selective activation of TVAs to human vocal sounds as compared to non-social naturally occurring sounds (Belin et al., 2004), such as car's noise, rustle of leaves, and well-matched acoustic controls, suggests a special contribution of these cortical areas for the processing of socio-relevant vocal information, such as the human vocalizations and speech (Shultz, Vouloumanos, & Pelphrey, 2012).

In the subsequent sections, the brief analysis of the three types of vocal information processing in typically developing individuals is performed.

1.1.1 Vocal Affect Information

Well before the emergence of language, nonverbal vocal expressions served as the main auditory signal for social interaction and communication (Belin et al., 2004). Through vocalizations (e.g., scream, crying, and laugh), and more specifically through modulations of the tone of voice, it is possible to communicate emotional states, which unlike face cues are relatively independent of the speaker's presence (Liebenthal, Silbersweig, & Stern, 2016). The tone of voice is acoustically perceived as pitch (highness or lowness of a voice), in which high pitch is often correlated with positive basic emotional valence (e.g., happiness), and low pitch is more likely to signal sadness (Lattner, Meyer, & Friederici, 2005). The decoding of such information is crucial during social interaction, and includes the analysis and integration of multiple acoustic cues in a coherent percept, in order to accurately infer the speaker's emotional state and to behave accordingly (Gebauer et al., 2014).

This process starts very early in development, when a specific prosodic register is directed to infants during adult-child interaction and communication. Indeed, when speaking to infants, adults spontaneously use a prosodically distinctive speech register: infant-directed speech or *motherese*, in which exaggerated intonation is observed (Fernald, 1985). This specific speech register is characterized by higher voice pitch, wider pitch range, hyper articulated vowels, longer pauses, and slower tempo, which is linguistically simpler and includes redundant utterances and isolated words (Fernald, 1985; Saint-Georges et al., 2013). The intonational emphasis that it conveys serves as a reliable cue for the listener about a speaker's communicative intentions, such as attention bid, approval, irony, sarcasm (Nazzi & Ramus, 2003; Saint-Georges et al., 2013), decoding of which is essential for developing theory of mind, and intersubjectivity (Saint-Georges et al., 2013).

Interestingly, the selective infant's responsiveness to infant-directed speech seems to rely on a more general preference - for positive affect in speech, i.e., infants prefer to attend to the adult-adult communication than to infant-directed speech if it contains more positive affect (Saint-Georges et al., 2013). Moreover, because infants' auditory sensitivity and frequency discrimination abilities are better in the region of 500 Hz than in the region of 100 Hz (Fernald, 1985) of the vocal sounds, higher pitch voices may

have some perceptual advantage i.e., fundamental frequency characteristics may modulate affect-based preferences (Saint-Georges et al., 2013). In fact, when comparing with amplitude and duration acoustic patterns, the auditory preference for fundamental frequency modulations is observed in typically developing children when decoding human vocalizations (Saint-Georges et al., 2013).

In typically developing individuals, the processing of vocal affect information involves right lateralized brain regions (Belin et al., 2004), such as the anterior portions of the superior temporal lobe, as well as the inferior frontal gyrus, together with the orbitofrontal cortex (see Appendix) (Schirmer & Kotz, 2006). However, depending on the emotional valence, the positive or negative value of the stimuli, slightly separate brain regions are activated. For example, whereas angry prosody elicits the activation of the right temporal cortex, happy prosody elicits an increased response of the right inferior frontal cortex even in 7 months-old infants (Grossmann et al., 2010). The recruitment of subcortical regions, such as the amygdala and insula, is also observed during the processing of emotional tone of voice, more specifically when the salience of the stimuli, that is, the perceptual relevance of the stimulus is analyzed (Leitman et al., 2016). Both left and right hemisphere amygdalae are activated when analysing the salience of non-verbal vocalizations (Fruhholz, Trost, & Kotz, 2016), as well as affective speech prosody (Liebenthal et al., 2016). Indeed, both hemispheres are sensitive to emotion cues (Ethofer et al., 2012), but left hemisphere predominance is manly observed during high frequency variations, characteristic of phonemic information (Liebenthal et al., 2016).

The paralinguistic aspects of vocal affect information processing seem to be lateralized to the right hemisphere already in 4 year-old infants (Homae et al., 2007; Wartenburger et al., 2007). The neuroimaging findings are corroborated with behavioral evidence which indicates that the recognition of the emotional tone of voice constitutes a difficult task for an infant as young as 3 years of life (Yoshimatsu, Umino, & Dammeyer, 2016). Thus, perceptual tuning to the emotional tone of the voice very early in development, as well as the cortical specialization of voice-sensitive regions, is crucial for an infant to acquire effective communication and interaction abilities.

1.1.2 Speech Information

Speech is an evolutionarily new cognitive ability in humans, and represents a particularly complex and abstract use of voice (Belin et al., 2004) to allow humans to exchange thoughts, beliefs and desires through mental representation of semantic concepts (de Villiers & de Villiers, 2014). It is a continuous acoustic signal that adult listeners segment into meaningful linguistic units (Christophe et al., 2003), by retrieving the acoustic sound patterns of words, and connecting them to the lexical representations stored in the mental lexicon (Nazzi & Ramus, 2003). However, newborns start with no phonological or semantic knowledge (Nazzi & Ramus, 2003), which learning requires the identification of meaningful word forms from the dynamic auditory vocal input. The parsing of the continuous speech signal is a complex procedure for an infant due to the frequent absence of clear acoustic markers at word boundaries to signal meaningful speech information, e.g., silent pauses, because words typically do not occur in isolation (Christophe et al., 2003; Gervain & Mehler, 2010; Gervain & Werker, 2008).

Despite this difficult task, the infant's early auditory discriminative is striking, once newborns with 3 days of life already show preference for the sounds of their native language (Speer & Ito, 2009). Moreover, infants turn rapidly to detect the rhythmical and intonational properties of the native language around 2 months (Brooks & Kempe, 2012), activating the same left-lateralized language related brain regions as in adults already at 3 months of life age (Dehaene-Lambertz & Dehaene, 2002). The auditory tuning to prosodic cues such as pitch contour, specific intensity, and duration (Wartenburger et al., 2007), seems to be a key perceptual ability for an infant to extract critical sound features from complex speech signal and build categorical phonetic representations (Jacobsen, Schröger, & Alter, 2004). Furthermore, the prosodic properties, such as intonation, stress and rhythm tend to co-occur with word boundaries to the voice (Christophe et al., 2003; Gervain & Mehler, 2010; de Diego-Balaguer, Martinez-Alvarez, & Pons, 2016), suggesting the importance of attentional orienting towards this acoustic cues in order to segment meaningful linguistic units within a continuous speech signal (Jusczyk, 1999).

As mentioned previously in section 1.1.1, infant-directed speech highlights the linguistic units within an utterance thought exaggerated pitch modulations – prosody

(Ethofer et al., 2012), turning the parsing task of a continuous speech signal easier for an infant who is learning the native language. The early perceptual tuning to pitch modulations seems to be important not only for the decoding of suprasegmental parameters of emotional vocal stimuli but also for speech-relative information. The decoding of such acoustic parameters is important for the development of narrowed perception to the native sounds of speech, which occurs around 11 months of age (Kuhl, 2004). More specifically, the automatic extraction of fast spectral changes of fundamental frequency and formants, e.g., resonant frequencies F1, F2 (Pisanski et al., 2016) is crucial for vowel perception (Swanepoel, Oosthuizen, & Hanekom, 2012), from which the auditory system builds up phonemic representations (Jacobsen et al., 2004).

At the brain level, the posterior superior temporal gyrus seems to serve an auditory-motor template interface (Hickok & Poeppel, 2007), tuned to fast temporal auditory features such as phonemes (Liebenthal et al., 2016). The activation of this temporal brain region, functionally connected with the pars opercularis of the inferior frontal gyrus (Friederici, 2012), is important for lexical phonological processing, that is, the mapping of phonemic representation onto lexical entries (Lukatela & Turvey, 1991).

In sum, the perception of speech information depends crucially on the ability to attend automatically to spoken stimuli (Vouloumanos et al., 2001), to segment the meaningful units from the continuous auditory stream (Vouloumanos & Werker, 2007), and to learn their relation to each other, deriving words and rules of native language (Mueller, Friederici, & Mannel, 2012). However, the communication skill is more than the recovery of segmental representation into meaningful linguistic units, it also comprises the ability to understand the intonation in order to infer indirect meaning conveyed by irony, sarcasm, and intentionality frequently used by speakers in order to act appropriately when socially interacting (Christopheet al., 2003; Brooks & Kempe, 2012). Thus, segmental as well as suprasegmental or prosodic features should be correctly perceived to ensure effective social communication and interaction.

1.1.3 Identity Information

As we may conclude from the previous sections, the human voice provides a wealth of linguistic and paralinguistic social information, including who is speaking (Abrams et al., 2016). The ability to recognize a specific individual within a social context is the foundation of social cognition (Zilbovicius et al., 2013). Because of the peculiar configuration of the vocal tract of each speaker, and more specifically of the set of laryngeal and supralaryngeal parameters (Mendoza et al., 2016), we are able to learn the characteristics of voices that are unique to each speaker, and thus recognize speakers whom we have heard before (Cutler, 2012).

Resonant frequencies (F1, F2…) are not only relevant for the identification of critical phoneme information important for speech comprehension, but also for the storage of the characteristic voice quality or timbre of a speaker (Jacobsen et al., 2004). The number of vocal fold vibrations determines the perceived pitch of a voice (Lattner et al., 2005). Listeners often associate lower frequencies with stereotypically male traits (Pisanski et al., 2016). Interestingly, lower pitch encodes not only the male gender; its perception is also important for emotional recognition, as described in the previous section of vocal affect information. For example, sad emotions tend to be communicated with a lower tone of voice (Lattner et al., 2005). Neuroimaging studies show that male formant configurations activate the pars triangularis of the left inferior frontal gyrus, whereas female voices seem to recruit more the supra-temporal plane in the right hemisphere, as well as the insula (Lattner et al., 2005). The supra-temporal plane, localized within the auditory cortex, plays a key role for the processing of complex acoustic properties (Tremblay, Baroni, & Hasson, 2013), e.g., speech, which functioning is crucial for the perception of formant frequencies (Formisano et al., 2008). The infant's preference over female voices could be explained by the higher perceptual salience of female timbre configuration for infant's auditory system, aroused by high-pitched voices (Lattner et al., 2005).

Within the right hemisphere, different brain regions are recruited during voice recognition and discrimination processes, where increased neural activity of the right anterior superior temporal sulcus (Belin et al., 2004; Bethmann, Scheich, & Brechmann, 2012; Kriegstein & Giraud, 2004) is observed during the recognition of famous or

familiar people (Van Lancker et al., 1988). On the other hand, the discrimination of unfamiliar voices seems to depend on the right posterior superior temporal sulcus, suggesting its key role in non-verbal complex temporal acoustic processing (Kriegstein & Giraud, 2004). The posterior portion of the voice sensitive regions seems to be essential for the storage of new incoming information from the auditory input in general, which could be a stranger's voice or phonemic level information (Buckingham, Hickok, & Humphries, 2001). Moreover, this brain region seems to be responsible for multisensory integration of noisy auditory and visual speech (Ozker et al., 2017), as well as for neural speech representation, as a prerequisite for hearing sounds such as speech (Möttönen et al., 2006).

The anterior temporal lobes seem to be crucial for accessing social knowledge in general, and are involved not only when retrieving memory traces associated with familiar voices (Fujii et al., 1990; Lattner et al., 2005), but also when recognizing other's thoughts and feelings, and when predicting their reactions (Bethmann et al., 2012). Bilateral activation of these areas is observed when testing the subjects' theory of mind tasks, as well as when contrasting between abstract social and non-social concepts (honorable vs. nutritious) (Zahn et al., 2007). However, hemispheric asymmetry is observed: the left anterior portion of the superior temporal gyrus is mostly activated to determine whether a heard word form can be semantically interpreted (Skeide & Friederici, 2016), the right anterior portion of superior temporal gyrus is strongly recruited during affective prosodic information processing instead (Schirmer & Kotz, 2006). Lesions in the anterior temporal lobes cause person recognition deficits (Bethmann et al., 2012), including memory about people's traits, their names and biography (Olson et al., 2013).

Altogether, this specific brain region seems to be crucial for social interaction skills development that include not only the ability to decode the prosodic information from a voice, but also to attribute who is conveying that information.

2. AIMS OF THE STUDY

The aim of the present project is to clarify how different types of vocal information are processed by individuals with ASD compared to typically developing individuals. Based on a cognitive science approach, the current systematic review integrates interdisciplinary evidence about voice and speech processing in ASD coming from different scientific fields: psychology, linguistics and neuroscience. The main goal is to understand if vocal information processing occurs differently in individuals with ASD compared to typically developing individuals, and which type of information conveyed by the voice causes main perceptual difficulties for its processing.

In order to collect all relevant scientific studies and provide behavioral and neurobiological outcomes, the present systematic review follows specific stages described by PRISMA statement, as well as the model of voice perception proposed by Belin and collaborators (2004). To our best knowledge, there is no systematic review analysing auditory stimuli taking into account the three types of information that humans may convey through the voice. The three different types of vocal information are processed by partially dissociated cerebral pathways, which constitute a good opportunity to detect which of the pathways may be mostly impaired in individuals with ASD. Based on neuroimaging evidence about the reduced activity of voice-sensitive brain areas and the reduced automatic orienting towards vocal sounds, the hypothesis of a impaired recognition of emotional value from socially relevant auditory stimuli in ASD population is analysed.

On the other hand, a low level of endogenous oxytocin in individuals with ASD represents one of the possible biological evidence for the reduced recognition of emotional value from the vocal sounds. Therefore, the effect of intranasal oxytocin administration is analysed in order to understand a possible improvement in the processing of social emotions by individuals with ASD. In the end, the explanatory hypothesis of ASD – Theory of Mind – is briefly discussed taking into account the results about vocal information processing from behavioral and neuroimaging studies included in the present systematic review.

3. METHODOLOGY

In order to create an accurate systematic review about different types of vocal information processing in ASD, several methodological approaches and steps were followed. Scientific evidences coming from studies with behavioral measures, electrophysiological and neuroimaging techniques were collected to shed light on how individuals with ASD differ from typically developing individuals when processing vocal information.

The spatial resolution of neuroimaging techniques, more specifically of fMRI, is necessary to observe which brain regions are activated when auditory information reaches the central nervous system (Goense, Bohraus, & Logothetis, 2016). Therefore, critical information about *where* the specific type of vocal information was processed is available. Neuroimaging data may be complemented by the electrophysiological evidence, which provide the exact temporal resolution, in order of milliseconds, for a clear understanding of *when* different stages of voice information take place (Luck, 2005). Complementarily, behavioral studies provide additional information about *how* individuals with ASD perceive different types of vocal information, by evaluating and integrating both verbal and nonverbal cues. For example, while neuroimaging studies may reveal similar brain regions activation in ASD compared to their typically developing controls, behavioral studies may show differences (e.g., in recognition accuracy or reaction times) in ASD.

Indeed, the integration of different methodological techniques allows a more comprehensive view of how individuals with ASD differ from their typically developing controls, when asked to process speech, affect and identity vocal information.

3.1 PRISMA Guidelines

A systematic search of the literature outcomes was performed, in order to collect all relevant studies about vocal information processing in individuals with ASD, as well as the possible effect of exogenous oxytocin effect over emotional stimuli processing.

For this purpose, the PRISMA guidelines were followed with its specific phases of a qualified systematic review. The selected electronic databases for the present review were the PubMed (National Library of Medicine National Institutes of Health) as well as the PsycINFO (American Psychological Association) from which 41 scientific article's outcomes reviewed in full were analyzed and integrated. The heterogeneous nature of ASD, and more specifically the lack of a clear definition of what the spectrum is, does not allow the analysis of restrictive samples. Indeed, the spectrum includes not only high functioning Autism, but also the Asperger Syndrome, and pervasive developmental disorder not otherwise specified (American Psychiatric Association, 2013). However, in order to achieve more homogeneity in the reviewed results across studies, only those studies examining individuals with high functioning autism and with Asperger Syndrome were included.

Furthermore, other inclusion criteria were: published studies written in English language with quasi-experimental designs randomized controlled trials and qualitative research where the number of participants was not fewer than 10 individuals. In order to extract all the information from a study, a table with authors, publication year, number of participants, age range, study design, task and outcomes, was made. Among specific stages of a systematic review postulated by PRISMA guidelines, the number of reports identified through the main databases searching is of 118 articles from PubMed database and 42 additional records identified through from PsycINFO source (see Figure 1). The collected scientific studies are the result of the following key-words introduced in the databases: "(oxytocin OR MRI OR neuroimaging OR BOLD OR EEG OR ERP) AND (prosody OR auditory OR voice OR vocal) AND autism.

Further in the screening phase, after all duplicates have been removed, a total of 130 studies were collected. However, after reading all papers' abstracts, 75 studies were excluded where the exclusion criteria were small population size, heterogeneous

population, or visual stimuli. At the end, only 41 studies were included in order to integrate evidence coming from different methodological measures.

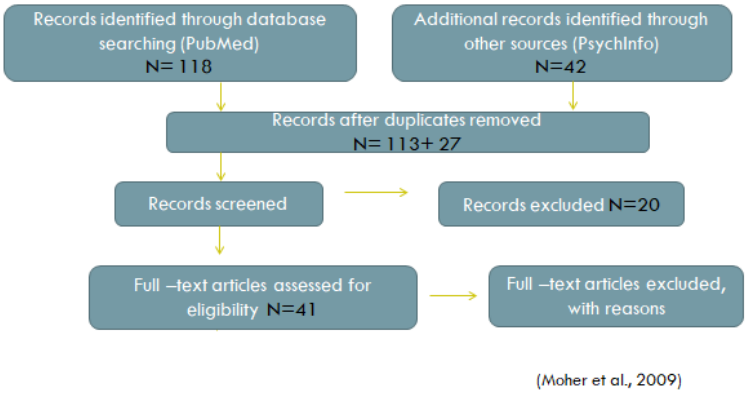

Figure 1: Following the PRISMA statement, a total of 118 scientific studies were found through the PubMed database. Moreover, additional 42 studies from PsychInfo database were obtained. After a careful analysis of each article's abstract, a total of 55 studies were read. However, 15 studies were excluded due to heterogeneous samples, visual stimuli or a small number of participants.

3.2 Voice Perception Model

The explanatory model of voice perception proposed by Belin and collaborators (2004) was followed, which proposes partially dissociated neural pathways for the three types of information that humans may convey through the voice (see Figure 2). Human vocal sounds are rich in different types of information, including the verbal message that contains linguistic information, as well as paralinguistic information from which the listener may not only decode cues about the emotional state of the speaker, but also about other identity characteristics, such as his/her gender and approximate age (Belin et al., 2004; Shultz et al., 2012). Therefore, the analysis of each type of information is important for a better comprehension of which vocal information is perceptually more difficult to process by individuals with ASD.

Typically, an early activation of the primary auditory cortex is observed in response to acoustic parameters analysis of the vocal sounds as the first stage of voice information decoding (Ardila & Bernal, 2016; Litovsky & Hearing, 2015). After the low-level acoustic analysis of vocal stimuli in voice-sensitive temporal areas, including the

bilateral superior temporal sulcus/gyrus (STS/G), voice structural encoding takes place in order to categorize the voice input. When processing speech stimuli, a subsequent activation of left posterior portions of the superior temporal gyrus is observed during the initial analysis of phonemic information (Hullett et al., 2016), whereas the attribution of the semantic label to the perceived acoustic signal (Skeide & Friederici, 2016), and the comprehension of complex social concepts and mental states (Shultz et al., 2012), recruits predominantly the anterior portion of this cortical area.

On the other hand, the analysis of vocal affective information is mostly observed within right lateralized voice-sensitive areas, as well as temporo-medial regions, amygdala (see Appendix) and inferior frontal gyrus (Hullett et al., 2016; Schirmer & Kotz, 2006). When recognizing a familiar voice, the predominant activation of the right anterior portion of the STS/G is observed, suggesting the importance of this brain region for identity information processing in particular (Belin et al., 2004; Lancker, & Tartter, 1987). The dissociated neural pathways for the processing of each type of vocal information turn useful to clarify which type of vocal information – speech, affect, identity, is more impaired in individuals with ASD.

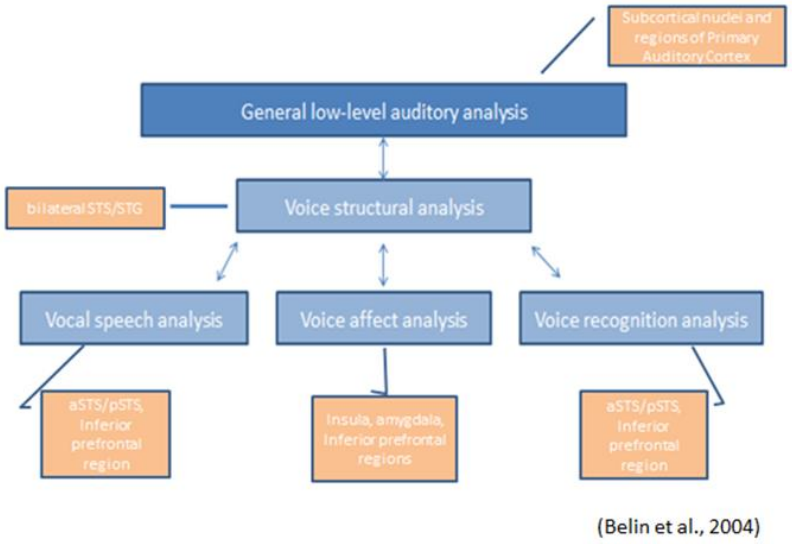

Figure 2: Three partially dissociated functional pathways for voice processing. After the acoustic analysis of the auditory signal, vocal speech, affect and identity information are processed by slightly separate brain regions.

4. RESULTS

4.1 ASD and Vocal Affect Information

Out of 41 selected empirical studies, 18 studies examining emotional prosody processing in ASD population across all age ranges were included in the present systematic review (see Table 1). Both nonverbal vocal information and speech information uttered with emotional intonation served as stimuli.

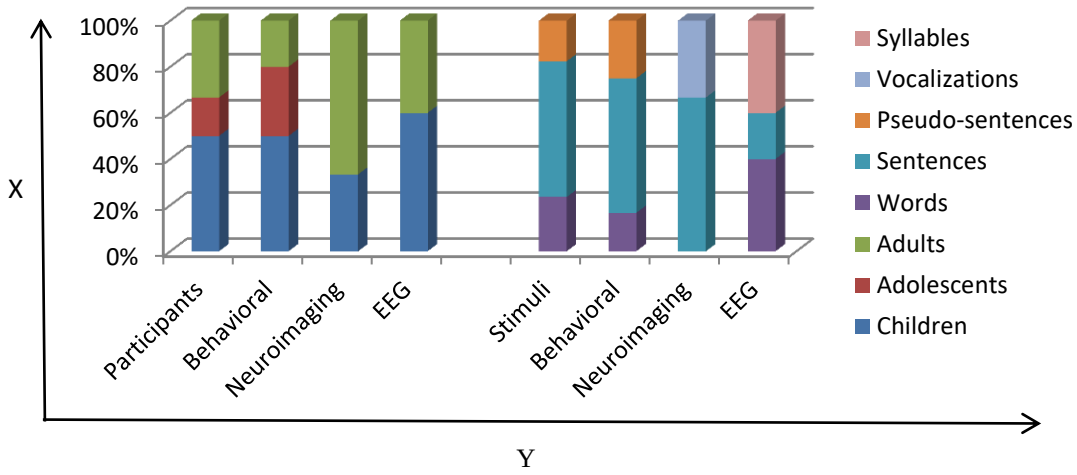

Table 1: From a total of 41 studies, 18 studies have examined the vocal affect information processing in individuals with ASD. From the X axis, it is possible to compare the age groups that were examined in the behavioral, neuroimaging and EEG studies, as well as the main type of stimuli used across all the studies and presented in the Y axis. For example, children with ASD are a major age group, representing 40% of studies (X axis), in which the recognition of emotional vocal sounds was examined. Nevertheless, this age group was predominant only in behavioral and EEG studies (Y axis), because neuroimaging studies have mostly examined vocal affect in ASD adults. From the Y axis we can conclude that the main type of stimuli used in the studies examining vocal affect information processing in individuals with ASD were sentences in behavioral and neuroimaging studies, and syllables in EEG studies.[1]

Behavioral studies (see Table 1.1) indicate lower mean recognition accuracy of emotional cues in children with ASD compared to typically developing controls when presented with semantically neutral sentences spoken with happy, sad, angry or scared intonation (Mazefsky & Oswald, 2007) The impaired processing of emotional prosody

[1] Among different nonverbal vocal and speech stimuli examined in vocal affect information processing studies, emotional vocalizations as well as pseudo-sentences uttered with emotional prosody were used. Vocalizations and pseudo-sentences are meaningless vocal stimuli that express an emotional intonation.

might be *specific* of certain types of emotions. For example, the recognition of basic emotions (happy, sad, afraid, angry, disgusted, and surprised) expressed within semantically neutral sentences is easier to recognize for children with ASD compared to complex emotions (i.e., interested, bored, excited, worried, disappointed, and frustrated) (Fridenson-Hayo et al., 2016). Nonetheless, children with ASD still perform at a less accurate and slower level compared to typically developing controls, suggesting impaired identification of socially complex mental states in ASD population (Fridenson-Hayo et al., 2016). Longitudinal studies show that 6-15 years-old ASD individuals present more correct responses as compared to 3, but not 5 years old typically developing children in emotion recognition, both when prosody is congruent or incongruent with the semantic content of speech, i.e., "positive prosody with positive semantic meaning/ negative prosody with positive semantic meaning" (Yoshimatsu et al., 2016). These results indicate that the ability to recognize emotional intonation embodied in speech stimuli is not established at the age of 5 years in children with ASD as in typically developing children. In older ASD groups, the ability to recognize emotions is more accurate when the emotional intonation is congruent with the semantic content of the sentence (Stewart et., 2013). However, adults with ASD persist as less accurate than their typically developing controls in trials of incongruent semantic and prosodic content, more specifically when the linguistic content is negative and the prosody is positive (Le-Sourne Bissau et al., 2013; Stewart et al., 2013). These results suggest that when there is concurrent semantic and emotional prosodic information, individuals with ASD have difficulty in ignoring the semantic meaning and to pay attention only to the affective tone of voice.

The recognition of emotional value seems to be more accurate, but slightly slower when children and adolescents with ASD are presented with low-pass filtered sentences or meaningless sentences with no recognizable semantic meaning (Grossman et al., 2010; Brennand et al., 2011). However, even in the absence of intelligible semantic meaning, the ASD group still shows lower accuracy in the recognition of happy prosody compared to typically developing controls (Baker et al., 2010). Altogether, a specific perceptual deficit in identifying positive valence is observed compared to the other emotion types, both in the presence of concurrent semantically incongruent contexts, as well as within meaningless sentences uttered with emotional prosody (Baker et al., 2010; Brennand et al., 2011; Le-Sourne Bissau et al., 2013).

Publication	Sample	Task	Results
Mazefsky & Oswald, 2008	AS n=16 (MA=11.47) HFA n=14 (26 male; MA=11.00)	ER in which participants were asked to identify the tone of voice from a sentence (happy, sad, angry or scared)	AS=HFA similar accuracy in perceiving high intensity tone of voice cues. AS<HFA significantly more difficulty with low intensity tone of voice
Baker et al., 2010	HFA n=19 (13 male; AR=10-14) TD n=19 (13 male; AR= 10-14)	ER in dichotic task from pseudo-sentenced uttered with angry, happy, sad and neutral intonation	HFA<TD poorer right-hemisphere ear effect during emotion presentation, and lower recognition of happiness
Grossman et al., 2010	HFA n=16 (AR= 7;6–17) TD n=15 (AR= 7;6–18)	APP in low-pass filtered and unfiltered sentences uttered with three emotions (happy, sad, and neutral) + Perception of linguistic prosody	HFA=TD similar categorization of sentences containing sad or happy emotions, as well as in the ability to perceptually disambiguate compound nouns from noun phrases using only prosodic cues
Brennand et al., 2011	AS n=15 (14 male; AR=10.5-19.3) TD n=15 (12 male; AR= 11-16.7)	ER from meaningless sentences uttered with angry (hot/cold), fearful (anxiety/panic fear), happy (happy/elated) and sad tone of voice (sad/desperation)	AS=TD similar results in the identification of basic emotions when instantiated on pseudo-sentences that do not have semantic content
Stewart et al., 2013	HFA n=11; TD n=14 (AR=17-39)	ER from vocalizations and sentences uttered with happy, fearful, surprised, angry tone of voice, which was emotionally congruent or incongruent with semantic meaning	HFA=TD similar emotional recognition from congruent context. HFA< TD poorer for the emotion recognition from incongruent context and vocalizations
Le Sourn-Bissauoui et al., 2013	HFA n= 26 (all male; AR=9.1-17.9) TD n=26 (all male; AR= 9.2-17.8)	ER from meaningless utterances conveying emotional prosody (positive or negative) embedded in a incongruent context (positive prosody + negative semantic meaning vs. neutral tone of voice	HFA<TD significantly lower prosody-based and higher linguistic-based responses in incongruent situation, when positive prosody was embedded in a incongruent context
Hsu & Xu, 2014	HFA = 10 (8 male; AR= 13;04–18;11) TD n=10 (8 male; AR=13.08-18.06)	ER from sentences uttered with emotionally 'neutral' voice and modulated by speech synthesiser to obtain breathy and pressed voice	HFA<TD less sensitive to manipulated voice quality. They seemed to rely mainly on formant shift ratio and pitch shift, and on pitch range to a less extent.
Globerson et al., 2014	HFA N=21 (23 male; AR= 20-40) TD n=32 (all male; AR= 23-39)	PD in which participant were asked if the tone was different from the previous one. ER in which asked to decide which emotion (happiness, sadness, anger and fear) was conveyed in the utterances. Pragmatic focus task in which there was accented word	HFA<TD poorer performance on vocal emotion recognition, but not on pragmatic prosody recognition or pitch discrimination.
Yoshimatsu et al., 2016	HFA n=35 (28 male; AR=6-15)TD n=39 (17 male; age= 5); n= 55 (29 male; age=4); n=64 (15 male, age=3)	ER from sentences in congruent context (positive prosody and positive semantic meaning) and incongruent context (positive prosody and negative linguistic meaning and vice-versa	HFA>3 years TD but lower than 5 years-olds TD in recognition of emotional prosody from congruent and incongruent context

Fridenson-Hayo et al., 2016	Israel HFA n= 20 (18 male; MA= 7.45); TD n=22 (19 male; MA= 7.50);Britain HFA n=16 (15 male; MA= 8.58); TD n=18 (13 male; MA=7.80) Sweden HFA n=19 (15 male; MA=6.97); TD n=18 (15 male; MA= 7.36)	ER from sentences with basic (happy, sad, afraid, angry, disgusted, surprised) and complex emotions (interested, bored, excited, worried, disappointed, frustrated, proud, ashamed, kind, unfriendly, joking, hurt).	HFA<TD showed lower scores on complex emotions recognition but with similar scores on basic emotions

Table 1.1: A detailed description of each behavioral study of the present systematic review examining vocal affect information in individuals with ASD. In total, difficulties in emotional value recognition from concurrent verbal content, delay in prosodic abilities development, as well as deficit in implicit processing of emotional tone of voice are observed in individuals with ASD. Legends: AS: Asperger Syndrome; HFA – High Functioning Autism; TD – Typically Developing individuals; ER – Emotion Recognion; APP – Affective Prosody Perception; PD - Pitch Discrimination; AR – Age Range; MA – Mean Age.

A total of 5 electrophysiological studies (see Table 1.2) examining vocal affect information processing in ASD were included. When presented with semantically neutral sentences spoken with a happy, sad, angry, or fearful emotional intonation, children with ASD present longer response latencies of M1n. The M1n is the neuromagnetic correlate of the N100 early ERP component originated within to primary auditory cortex and elicited when an unpredictable auditory stimulus is presented (Demopoulos et al., 2016). Longer latency indicates that the detection of the emotional stimulus does not occur at the same time interval as in typically developing children, which is delayed in children with ASD. However, no latency difference between groups is observed in the P3a component when presented with the repeated neutral word interrupted by a deviant scornful, commanding and sad intonation (Lindström et al., 2016).

The P3a component indicates the attentional orienting towards a novel stimulus that represents an infrequent sound pattern in an acoustic stream of repetitive (standard) sound sequences (Hruby & Marsalek, 2003). Shorter P1 component latency response is also observed in children with ASD when presented with meaningful syllables uttered with a steady pitch contour for standard stimuli and with a falling pitch for deviants (Yoshimura et al., 2016). The P1 component is generated within the primary auditory

cortex and thalamus (Sharma et al., 2015) indicating an early stage of auditory processing. Shorter P1 latency response in children with ASD suggests early activation of relevant cortical areas for low-level acoustic information processing, and early perceptual tuning to each acoustic parameter, which may interfere with the evaluation of the emotional value of vocal input. Altogether, these results suggest that although there is an attentional allocation towards novel sounds in ASD group, the recognition of its emotional value is delayed.

The impaired assignment of emotional value embodied in linguistic information is also reflected in the Late Positive Component, whose amplitude is diminished in adults with ASD when presented with emotionally positive or negative vs. neutral words (Lartseva, et al., 2014). The mismatch negativity is another ERP component elicited when an infrequent stimulus is presented. However, its reduced amplitude is observed in adults with ASD for meaningless syllables spoken with angry and happy emotional prosody vs. acoustically matched nonvocal sounds (Fan & Cheng, 2014).

Publication	Sample	Task	Results
Fan & Cheng, 2014	HFA n=20 (19 male; AR= 18-29) TD n=20 (19 male; AR=18-29)	PL to meaningless syllables /dada/ uttered with angry or happy tone of voice vs. acoustically matched non-vocal sounds	HFA failed to exhibit differentiation between angry and happy tone of voice vs. acoustically matched non-vocal sounds. HFA<TD reduced P3a to emotional voices
Lartseva et al., 2014	HFA n=21 (14 males) TD n=21 (15 males; AR=18-36)	LD in which the participants were instructed to read the letter string and respond whether it was an existing or not existing (pseudo) word with neutral, positive, and negative valence	TD=HFA N400 amplitude was significantly lower for low-frequency words in both groups. ASD failed to show a typical late positive component for emotion vs. neutral words
Demopoulos et al., 2016	HFA n=37 (AR=5-18) TD n=15 (AR= 5-18)	DANVA-2 test the ability to identify sad, angry, or fearful emotional content in neutral statement; (2) RAP - single tone 500 or 1000Hz presented in rapid succession	HFA<TD longer M1 latency to emotional vs. neutral stimuli and impairments in rapid auditory processing in the left hemisphere
Yoshimura et al., 2016.	HFA n= 35 (27 male; AR= 38-111 months) TD n=35 (27 male; AR= 32-121)	PL to syllable /ne/ Uttered with a steady pitch contour for the SD, and with a falling pitch under the DV condition.	TD>HFA U-shaped growth curve for the P1m dipole intensity in the left hemisphere and more diversified age-related distribution of auditory brain responses in ASD. HFA<TD shorter P1m latency bilaterally
Lindstrom et al., 2016	HFA n=10 (9 male; AR= 8-6-12.2) TD n=13 (12 male; AR= 7.5-11.8)	PL to words uttered uttered neutrally for SD stimuli and with scornful, commanding or sad voice for DV condition	HFA<TD smaller MMN and P3a amplitude for the scornful DV stimuli

Table 1.2: A detailed description of each electrophysiological study of the present systematic review examining vocal affect information in individuals with ASD. In total, impaired identification and orienting to the complex emotions, difficulties in recognizing emotional value in speech, as well as difficulties in inhibiting concurrent verbal information and focus on emotional intonation instead are the main perceptual deficits with vocal affect information processing in ASD. Legends: PL –Passive Listening; LD – Lexical Decision; RAP – Rapid Auditory Processing; DANVA-2 – The Diagnostic Analysis of Nonverbal Behavior; FG – Frontal Gyrus; SPL – Superior Parietal Lobe; SD – Standard stimuli; DV – Deviant stimuli; STG – Superior Temporal Gyrus, TP – Temporal Pole.

Only three neuroimaging studies (see Table 1.3) were incorporated in the present systematic review, examining how ASD participants process emotional voice information at the neural level. When presented with emotionally sad vocalizations vs. neutral voice, infants with low-risk for developing ASD showed activation of the left superior and the right inferior frontal gyrus, whereas a small cluster within the right cingulate gyrus was activated in high-risk infant's brain (Blasi et al., 2015). The absence of the typically rightward inferior frontal gyrus activation in infants at risk suggests disrupted recognition of emotional value from acoustic stimuli. Furthermore, whereas low-risk infants showed preference for listening to voices and bilateral middle and superior temporal regions activation, infants at risk showed a tendency to respond to non-vocal stimuli showing major brain activation within the right inferior parietal lobule (Blasi et al., 2015). Therefore, atypical voice processing seems to be present as early as at 5 month-old infants at risk for developing ASD (Blasi et al., 2015), a critical developmental period for the cerebral specialization for voice processing in typically developing infants (Grossmann et al., 2010).

Interestingly, in older ASD groups, activation of bilateral superior/medial temporal gyri, right inferior frontal gyrus, and orbitofrontal regions is observed when listening to semantically neutral sentences spoken with happy, sad and neutral prosody (Gebauer et al., 2014). Contrariwise, when presented with complex vs. basic emotions, reduced activity was observed in the bilateral superior temporal sulcus/gyrus and in the right amygdala in adults with ASD (Rosenblau et al., 2016). Thus, the activation of voice-sensitive areas is stronger when processing basic emotions, suggesting difficulty in interpreting higher-order mental states. Moreover, the activation of similar brain regions as in controls during the processing of emotional prosody may result from the attentional orienting for the expected emotional value present in the verbal message. Both studies (Gebauer et al., 2014; Rosenblau et al., 2016) have used explicit

instructions to pay attention to the emotion and not to the semantic meaning of the voice.

Publication	Sample	Task	Results
Gebauer et al., 2014	HFA n=19 (17 males; AR=20-36) TD n=20 (18 male; AR=19-41)	ER task from sentences uttered with happy, sad vs. neutral tone of voice	ASD > TD increased activation in middle/superior FG, SPL during processing of happy prosody, as well as of the right caudate in response to emotional vs. neutral prosody, while TD displayed increased activation of the left precentral/rolandic operculum.
Blasi et al., 2015	HR n= 15 (10 male; MA=147 +/- 25 days) LR n=18 (7 male; MA= 154 +/-26 days)	PL task in which emotional vs. neutral vocalization and environmental non-voice sounds were presented	HR did not activate de superior and medial temporal gyri for neutral vs. non-vocal sounds. HR did not show activation in the left SFG and right IFG for emotional vs. neutral vocalizations.
Rosenblau et al., 2016	HFA n= 27 (18 male, AR=19–47) TD n=22 (16 male, AR=20–46)	ER from sentences uttered with neutral vs. basic emotions (happy, surprised, fearful, sad, disgusted and angry) matched for valence and arousal with six complex emotions (jealous, grateful, contemptuous, shocked, concerned, disappointed)	HFA=TD increased activity of the STS and IFG during explicit evaluation of emotional prosody. HFA<TD reduced activity in bilateral STG, TP and right STS for complex vs. basic emotions.

Table 1.3: A detailed description of each neuroimaging study of the present systematic review examining vocal affect information in individuals with ASD. In total, infants at risk for developing autism show no specialization for human voices processing in the right temporal and medial frontal regions, as well as increased effort in evaluating the emotional significance of vocal stimuli is observed in children with ASD. However, when orienting their attention, the emotion recognition is similar to typically developing children. HR – High Risk infants; LR – Low Risk infants; STS – Superior Temporal Sulcus; IFG – Inferior Frontal Gyrus.

The recognition of the emotional quality of a voice is essential to infer the emotional state of the speaker and act accordingly during social interaction and conversation. Most of the studies examining the affect vocal information processing in individuals with ASD indicate slower reaction times towards emotional vs. neutral words (Lartseva et al., 2014; Rosenblau et al., 2016). Moreover, individuals with ASD are less accurate in recognizing complex emotional states vs. basic emotions (Rosenblau et al., 2016), showing a tendency to rate emotional stimuli as less emotionally intense as compared to typically developing controls (less happy or less sad) (Gebauer et al., 2014). Interestingly, the recognition accuracy and correct judgments of happy prosody among other emotional pitch contours seems to be associated with worse performance in

individuals with ASD (Baker et al., 2010; Le Sourne-Bissau et al., 2013; Brennand et al., 2011, Wang & Tsao, 2015).

The electrophysiological studies indicate the preserved early recognition of emotional value in speech stimuli. The disrupted recognition of emotional value is observed through the reduced activation of voice-sensitive areas in infants at risk for developing autism when presented with emotional vs. neutral vocalizations. Nonetheless, shorter P3a latency response in ASD group suggests early attentional allocation towards emotional stimuli, but which seem to be less salient for ASD children who show reduced MMN amplitude to speech stimuli uttered with emotional intonation compared to non-speech vocal sounds. Early attentional allocation towards emotional stimuli, specifically when individuals with ASD are explicitly instructed to pay attention to the emotion and not to the semantic meaning, seem to modulate voice-sensitive cortical activity in ASD, who show similar response of the bilateral superior/medial temporal gyri as control group.

4.2 ASD and Speech Information

After a careful selection of scientific articles examining speech information processing in individuals with ASD, a total of 17 studies were included in the present systematic review. The behavior evidence prevail the number of electrophysiological and neuroimaging findings (see Table 2). The majority of studies used children and adolescents, in which only three out of seventeen studies used adult ASD population.

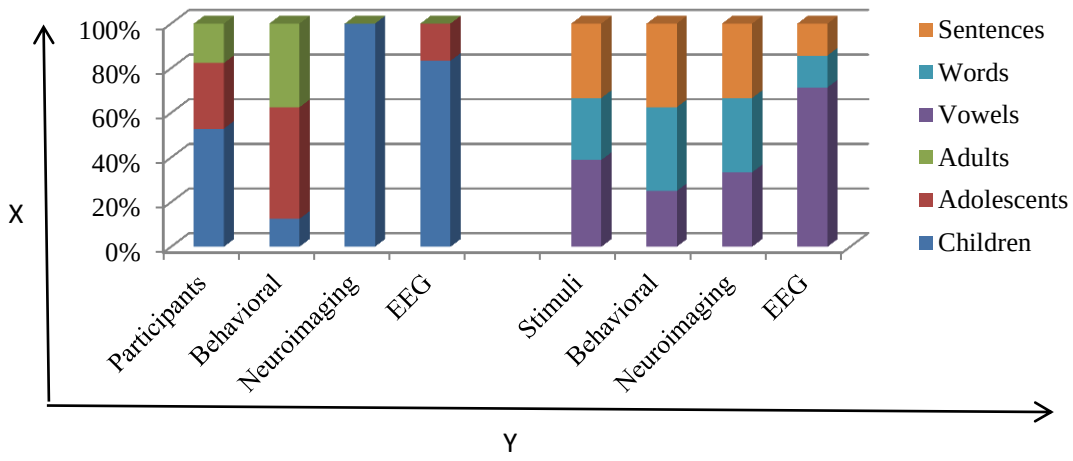

Table 2: From a total of 41 studies, 17 studies have examined the speech information processing in individuals with ASD. From the X axis, it is possible to conclude that most studies have used children and adolescents with ASD, hence only behavioral studies (Y axis) have examined ASD adults. Although there is an apparent regular distribution of the stimuli used (vowels, words and sentences), its distribution among behavioral, neuroimaging and EEG studies is not uniform. For example, in the EEG studies the main speech stimuli examined in children and adolescents with ASD included phonemic information. The most regular pattern of speech stimuli is observed within behavioral studies in which 20% of the studies have applied vowels, 40% have used words and another 40% of the behavioral studies have examined the processing of sentences.

When explicitly instructed to pay attention to sung vowels vs. musical instrument sounds, adults with ASD outperformed their typically developing controls showing faster reaction times (Lin et al., 2016). However, in the case of spoken vs. sung vowels, less correct responses were observed from individuals with ASD when asked to determine whether the first or the last phoneme was different from the other two (DePape et al., 2012). Behavioral results (see Table 2.1) suggest that when there is extended temporal information, individuals with ASD exhibit better perceptual

discrimination of speech segments. Moreover, when presented with simple (native) and complex (foreign) meter, individuals with ASD show a smaller difference in discrimination of different sound patterns (DePape et al., 2012), which indicates less specialization for native sounds of speech.

When examining the stress pattern perception, e.g., "HOTdog/hotDOG", both children with ASD and their typically developing controls were more accurate in the first-syllable predominant stress patterns than the last-syllable stress items, suggesting fast recognition of stress position and word meaning (Grossman et al., 2010). However, when presented with meaningless syllable template, e.g., ABBB or ABAA, adults with ASD show less accurate judgments about shared syllable stress (Kargas et al., 2015). The heterogeneous outcomes are also observed within linguistic accent information processing. Adolescents with ASD show more difficulties in identifying the salient linguistic unit signalized by accent information e.g., I want *chocolate* ice cream (Paul et al., 2005). Nonetheless, when explicitly instructed to focus their attention on the newly acquired meaning of a sentence, for example, "There are birds in the park", adults with ASD did not differ in recognizing accented words in the sentence as compared to typically developing controls (Globerson et al., 2014).

In typically developing individuals, the perception of speech in noise is facilitated in a pink noise condition (two stimuli with very different spectro-temporal features) compared to ripple stimuli, which spectrotemporal parameters are very dynamic and distracting, very similar and thus difficult to separate from speech (Groen et al., 2009). However, individuals with ASD are less accurate in recognizing the speech stimuli from concurrent background noise, and this difficulty increases with the dynamic changes of speech information (Groen et al., 2009). The general difficulty in discriminating speech stimuli from a noise signal is observed in both hyper and non-hypersensitive individuals with ASD, who show an adverse response to vocal sounds (Dunlop, Enticott, & Rajan, 2016). Both groups perform poorly in informational multi-talker babble noise, i.e., multiple people speaking at the same time, but with the same level of correct responses when perceiving speech from energetic speech-weighted noise, i.e., in which some acoustic parameter is increased or suppressed (Dunlop et al., 2016).

The enhanced perceptual sensitivity to auditory stimuli was recently highlighted as one more manifestation of restricted, repetitive patterns of behavior, interests, or activities in

individuals with ASD, such as hyper or hyporeactivity to sensory input, including adverse response to sounds (DSM-V, 2013). Although this criterion varies from individual to individual, the estimated prevalence is approximately 40% in ASD population (Dunlop, 2016). Moreover, the believed absolute auditory pitch discrimination ability in individuals with ASD has an irregular developmental pattern. The study of Mayer and collaborators (2016) using both words and analogue contour stimuli show that the percentage of correct scores for both stimuli was higher in ASD children and adolescents as compared to their controls, but lower when adult groups were compared (Mayer, Hannent, & Heaton, 2016). Whereas the discrimination ability of typically developing controls increased proportionally over age, a clear hypersensitivity during the early stages of development and its deterioration late in life was observed in individuals with ASD (Mayer et al., 2016)

Individuals with ASD show enhanced awareness of perceptual information in speech, making more correct responses when the task is focused on voice intonation compared to typically developing controls, and the same percent of correct responses when the task is focused on semantic content of speech (Järvinen-Pasley, Pasley, & Heaton, 2008). This study completes the findings of Mayer et al. (2016) which points to the superior pitch discrimination abilities in ASD children as compared to controls, and thus more attentional orienting towards the suprasegmental acoustic parameters, such as fundamental frequency.

Publication	Sample	Task	Results
Paul et al., 2005	AS n=27 (MA= 16.8) TD n=13 (MA= 16.7)	PP from sentences with three aspects of prosody: stress and intonation	AS<TD difficulty in understanding an appropriate stress patterns AS=TD no difficulty in distinguishing statements from questions, as well as child-directed from adult-directed speech (pragmatic/ affective perception of intonation)
Jarvinen-Pasley et al. 2008	AS n=28 (25 male; AR= 9.50–16.83) TD n= 28 (20 male; AR= 8.58–16.08)	Quasi open-ended paradigm to test semantic and perceptual (intonation) speech processing, in which the participants point to the picture that best matched the sentence heard previously	AS<TD provide weaker semantic interpretations of the stimuli, basing more on the intonational pitch contours (low-high-low, high-low-high) of the read sentences

Groen et al., 2009	HFA n=23 (19 male; AR=12-17) TD n= 23 (18 male; AR=12-17)	SiN with explicit instruction. After each presentation of concurrent masking background stimuli fragmented both temporally and spectrally (pink noise and moving ripple), the subjects immediately repeated what they heard	Smaller RT HFA<TD when perceiving words from amplitude modulated pink noise. HFA=TD in integration of auditory information fragments present in spectral dips of ripple sounds.
DePape et al., 2012	HFA n=27 TD n=27 (all male) (AR=11-18)	(1) Competing sentences test to ignoring simultaneous sentences; (2) PC to determine whether the first or last phoneme was different from the other two (3) Absolute pitch- indicating whether the two pitches were the same or different	HFA group showed evidence of filtering problems at the level of speech sentence, as well as lower specialization for native-language phonemic categories. Only 11% of HFA had absolute pitch, and the other 89% appeared to process pitch similarly as controls.
Mayer et al., 2016	HFA child n=14 (all male; AR= 6.11-14.9) TD n=14 (13 male; AR= 5.0-14.1) ASD adol. n=14 (13 male; AR= 9.8-16.5) TD n=14 (all male; AR= 12.0-16.9) ASD adults n= 19 (15 male; AR= 23.9-59.8) TD n=19 (14 male; AR= 25.1-52.8)	PD in words vs. non-speech stimuli. The participants were asked to indicate whether the two words in the pair were the same or a different pitch by pressing a button on a computer keyboard labelled ‘‘S’’ or ‘‘D’’.	In TD pitch discrimination abilities improved proportionately across age groups, whereas in HFA the discrimination ability is enhanced early in development and stable over time
Kargas et al., 2015	AS n=21 (18 male; MA= 30.3) TD n=21 (18 male; MA= 29.5)	SP of the position of syllable stress in the (e.g., **au**ditory and **da**ndelion, and words with second syllable stress: ca**pa**city and de**mo**cracy	AS<TD less sensitive in the detection of syllable stress. However, the performance on the syllable stress perception task varied considerably across individuals with AS.
Lin et al. 2016	HFA n=12 (9 male; MA= 27.5+/- 7.93) TD N=11 (9 male; MA= 27.27+/- 9.24)	RT and go/no-go task. Explicit instruction to attend auditory stimuli and respond as fast as possible only to a designated class of target stimuli: /i/, /a/ vowels or non-vocal sounds	HFA=TD significantly shorter RT's for vowels vs. non-vocal sounds when instructed to pay attention.
Dunlop et al., 2016	HFA n=16 (14 males; AR=20-52) TD n=34 (28 male; AR= 19-51)	SiN sentences with multi-talker babble and speech weighted noise	HFA<TD performed poorly in speech discrimination when presented in multi-talker babble

Table 2.1: A detailed description of each behavioral study examining speech information processing in individuals with ASD. In total, deficit to discriminate speech from background noise, to detect acoustic prominence (stress) in speech is observed in ASD, in which attentional system seems to modulate speech stimuli processing, once when attended, better accuracy and reaction times are observed. Legends: PP – Prosody Perception; SiN – Speech in Noise perception; PC – Phoneme categorization; SP – Stress Perception; PD – Pitch Discrimination.

Six of electrophysiological studies included in the present systematic review used children and adolescents as the target groups (see Table 2.2). Among different auditory event-related brain potentials (ERP), reduced P1 amplitude in response to native speech sounds vs. non-native phonemes was observed in children with ASD, with no differences in its latency (Ceponiene et al., 2003; Whitehouse & Bishop, 2008). The same pattern is observed even in infants at risk for developing ASD when processing native syllables (Seery et al., 2013). Dimished P1 amplitude suggests that the perceived speech information is less salient for children with ASD compared to typically developing controls. The topographical distribution of the auditory P1 component is similar in both hemispheres except for the left anterior temporal regions, in which larger amplitudes are typically observed compared to the right hemisphere (Key et al., 2005).

However, atypical rightward asymmetry observed in individuals with ASD (Floris et al., 2016), where infants at risk for developing ASD show faster responses for syllable processing in the right and not left hemisphere (Seery et al., 2013). Consequently, the reduced left hemisphere activity during speech processing, may contribute to the decreased P1 amplitude when compared with typically developing individuals. In typically developing individuals, the recruitment of the auditory areas is observed during native speech sounds processing as compared to non-native segments, which in contrast elicit greater activation of the motor areas (Kuhl, 2004). However, the absence of perceptual narrowing for the native phonemes in individuals with ASD may thus be explained by the dimished activation of the auditory areas (Samson et al., 2011) responsible for the structural encoding of the vocal stimuli.

The reduced automatic orienting towards speech signal is corroborated by studies probing the P3a component, which indicate diminished (Lepistö et al., 2005) or even absent (Ceponiene et al., 2003) amplitude of this component in children with ASD during the processing of native phonemes. The P3a amplitude is usually elicited when an unexpected sound is detected by the perceptual system (Polich, 2003). However, larger P3a peak amplitude is elicited when non-speech and complex tones are presented to ASD participants (Whitehouse & Bishop, 2008), suggesting an altered bias towards socially relevant information, and the lower salience of speech sounds for individuals with ASD.

Another ERP component frequently mentioned as atypically reduced during speech processing in children and adults with ASD is the mismatch negativity (MMN) ERP component. MMN originated within the auditory cortices, more specifically within the supratemporal planes of the temporal lobes (Pakarinen et al., 2014) reflects pre-attentive, not volitional detection of cues that deviate from a regular sound pattern, by comparing acoustic representation temporarily stored in the auditory sensory memory (Jacobsen et al., 2004). Interestingly, the reduced MMN peak amplitude is observed in infants with ASD for both words and pseudo-words (Ludlow et al., 2014), but enhanced during pitch change detection (Lepistö et al., 2008), suggesting enhanced perceptual salience of acoustic parameters of the auditory signal.

Interestingly, those children with ASD who prefer infant-directed speech show similar results of the typically developing controls, i.e. they exhibit a MMN response to a syllabic change during deviant speech syllable detection (Kuhl et al., 2005). On the other hand, those of children with ASD who prefer non-speech stimuli continue to show the abnormal MMN pattern (Kuhl et al., 2005). These findings indicate that when the infant's perceptual system is tuned to suprasegmental parameters of the voice, subsequent attentional orientation toward socially relevant stimuli occurs. Moreover, the early perceptual tuning to the mother's voice is an important key for the advanced diagnosis of ASD.

Publication	Sample	Task	Results
Ceponiené et al., 2003	HFA n= 9 (8 male; AR=6.3–12.4) TD n=10 (9 males; AR= 6.6–12.4)	PL to vowel /o/ vs. non-speech sounds	P1 amplitude HFA<TD. P3a response to complex tones vs. vowel changes in ASD.
Lepistö et al., 2005	HFA n=15 (13 male; (AR=7.3-11.10) TD n= 15 (13 male; age range= 7.5-11.11)	PL to vowels /a/, /o/ vs. non-speech sounds	MMN amplitude: HFA< TD speech vs. non-speech frontocentrally, but enlarged over parietal areas. P3a ERP component HFA< TD for phonemes changes
Lepistö et al. , 2008	HFA n=10 (9 male; AR=7.0–11.0) TD n=16 (15 male; AR= 6.11–10.10)	PL to vowels with different frequency modulation within 2 conditions: constant-feature – SD stimulus randomly replaced by DV stimuli and varying feature - constantly varying both SD and DV stimuli.	MMN amplitude: HFA<TD for standard phoneme changes in the varying-feature condition.
Whitehouse & Bishop, 2008	HFA n=15 (all male; AR=7.6–14.3) TD n=15 (11 male; AR= 7.0–14.3)	PL to speech conditions - /a/ SD vowel; /i/DV vowel and the novel sound was a 800 Hz complex tone. Non-speech conditions - the SD complex	P3 and P1 amplitudes: HFA<TD speech vs. complex tones. HFA=TD when attending to stimuli. P3 amplitude: HFA<TD novel tones after speech sounds,

		tone 500 Hz; DV 800 Hz complex tone and the novel sound was a vowel /i/.	but not to novel speech sounds after standard tones.
Seery et a., 2013	HR n= 14; LR n=12 observed over the first year of life 6, 9 and 12 months	PL to syllables: SD /ɖa/ native DV /ta/ and non-native DV/da/	HR<LR between 6 and 12 months did not display a typical left-hemisphere lateralization in response to the native speech sounds
Ludlow et al., 2014	HFA n=11 (all male; AR=11-16) TD n=11 (all male; AR= 11.11-15.8)	PL to words: SD baj/paj and DV bajp/ bajt/ pajt/ pajp	MMN amplitude HFA<TD for both words and pseudowords. N4 amplitude frontocentrally HFA<TD in response to SD speech stimuli but not to non-speech stimuli

Table 2.2: A detailed description of each electrophysiological studies examining speech information processing in individuals with ASD. In total, atypical lateralization to speech stimuli, impaired involuntary orienting to speech sounds, as well as reduced discrimination and recognition of speech sounds are observed in ASD.

Neuroimaging studies (see Table 2.3) reveal that when presented with speech-like stimuli with similar acoustic structure of consonant-vowel-consonant sequence vs. resting, children with ASD show less activation of the left middle temporal gyrus, as well as of the left precentral gyrus (Boddaert et al., 2004). The absence of the left hemisphere dominance in children with ASD, and further recruitment of additional cerebral areas outside the auditory cortex, including the bilateral posterior parietal cortex, the cerebellar hemispheres, and the brainstem (Boddaert et al., 2004), indicates difficulties in processing speech relevant units. Interestingly, the recruitment of parietal and cerebellar cortex is observed during the learning of a difficult second-language phonetic contrast in typically developing individuals (Graves et al., 2009), which supports the lack of tuning to speech native sounds in children with ASD.

When listening to familiar nouns and verbs vs. sung words, children with ASD show, once again, the absence of a left biased asymmetry within the inferior frontal gyrus as compared to typically developing children (Sharda et al., 2015). Interestingly, when the same words were presented in the sung condition, children with ASD alike the controls activated the superior temporal sulcus/gyrus and the medial temporal gyrus bilaterally (Sharda et al., 2015). However, when functional connectivity was compared in the sung context, children with ASD showed increased left inferior frontal gyrus connectivity with bilateral posterior temporal, right parieto-occipital, and left cerebellar regions, whereas their typically developing controls show increased connectivity within the left

inferior frontal cortex and right anterior insula (Sharda et al., 2015). These findings suggest that even when the temporal information is extended in the sung words context, the processing of linguistic information is still difficult for individuals with ASD.

The same pattern of response is observed during the passive listening to sentences, in which infants at risk for developing ASD show reduced response of the left superior temporal gyrus with right-lateralized responses of the anterior portion of temporal cortex (Eyler et al., 2012). The reduced activation of the left-lateralized brain regions may be explained by the recent results of Sharda and collaborators (2015), who found a reduced voxel-wise fractional anisotropy in a posterior region of the left Superior Longitudinal Fasciculus in infants at risk for developing ASD, which serves as a pathway connecting language-relevant brain regions (Sharda et al., 2015). The fronto-temporal connectivity is a crucial neuronal pathway for normal language functioning, and its impairments in individuals with ASD may explain the reduced left hemisphere activation and rightward asymmetry as a compensatory mechanism for speech information processing. The overcharge of the right hemisphere during both prosody and speech information processing may result in deficient comprehension of these types of information in individuals with ASD.

Publication	Sample	Task	Results
Boddaert et al., 2004	ASD n=11 (10 male; AR=4-10) TD n=6 (4 male; AR=3-9)	PL to speech-like stimuli with similar acoustic structure to consonant-vowel-consonant vs. resting-	ASD did not show left-biased asymmetry in response to speech-like stimuli. ASD<TD less activation of the left MTG to speech-like stimuli.
Eyler et al., 2012	HR n=40 (73% male, AR=12.5–47.6 months) TD n=40 (63% male; AR=12.3–45.3)	PL to segments of a children's story played forward and backward	HR<TD show reduced response of the left STG with right-lateralized temporal cortex response to speech; this defect worsens with age, becoming most severe in 3 and 4-year-olds.
Sharda et al., 2015	HFA n=24 (16 males; AR=6-16) TD n=22 (16 males AR= 6-16)	PL to familiar words ,sung words and non-speech sounds	HFA no leftward asymmetry in response to spoken words. TD=HFA showed similar bilateral temporal network STS, STG, and MTG in sung vs. spoken.

Table 2.3: A detailed description of each neuroimaging studies examining speech information processing in individuals with ASD. In total, all studies indicate right lateralization abnormalities, where right-lateralized brain regions are activated during speech tasks in ASD. Legends: MTG – Medial Temporal Gyrus.

In total, abnormal speech stimuli processing with right lateralization asymmetry seems to be a robust finding across all the neuroimaging studies. These studies indicate that those stimuli containing critical speech information acoustic parameters activate different neural pathways in individual's brain with ASD, suggesting altered perception system for speech information processing. The main difficulties with speech information processing seem to result in contexts where there is a concurrent background noise signal or prosodically accented linguistic units. The inhibition of task irrelevant information and perceptual tuning to target voice stimuli do not occur in individuals with ASD, who present bias towards the processing of each acoustic characteristic of the sound.

4.3 ASD and Identity Information

Fewer studies have examined the identity information processing in individuals with ASD, where only two behavioral and two neuroimaging studies were conveyed, with no electrophysiological evidence (see Table 3). Moreover, only adult individuals with ASD have participated in the studies, which makes impossible the comparison between younger and adults with ASD, and thus the developmental trajectory of identity information processing.

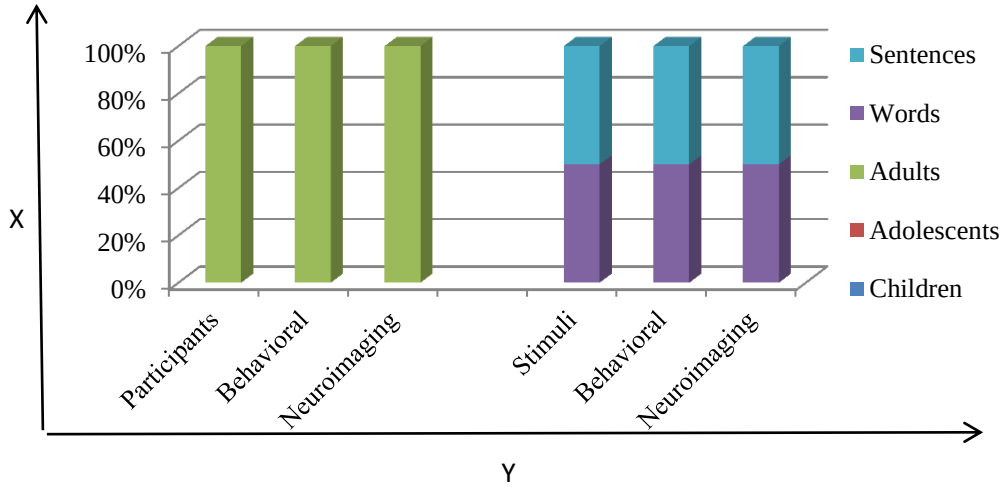

Table 3: From a total of 41 studies, only 4 studies have examined the identity information processing in individuals with ASD. Moreover, no children or adolescent studies were conducted, so the present results

only shed light on adult's ability to recognize and discriminate voices. Relative to the used stimuli, both words and sentences were presented uttered with familiar or unfamiliar voices.

The behavioral studies (see Table 3.1) included in the present systematic review used adults with ASD as the target experimental group and speech information produced by different speakers. Adults with ASD showed less accurate responses in discriminating whether two consecutive sentences were spoken by the same or different speakers (Schelinski, Roswandowitz, & von Kriegstein, 2016). On the other hand, in tasks involving a decision whether the heard voice was famous or not, ASD participants were at the same level as TD controls (Schelinski, Roswandowitz, & von Kriegstein, 2016) or even better when discriminating previously trained voices from untrained voices using words as stimuli (Lin et al., 2015). Possibly, the main difficulty lies on the discrimination of unfamiliar voices (Schelinski, Roswandowitz, & von Kriegstein, 2016), as well as of the familiar voices when the temporal information becomes larger, such as in sentences.

Publication	Sample	Task	Results
Lin et al., 2015	HFA n=14 (11 male; AR=20-47) TD n=14 (11 male; AR=20-43)	(1) GD - identify male/female voice (2) VIR - name the speaker presented on the screen who had produced target word (3) Familiarity test	HFA>TD in familiarity test when asked to discriminate the previously trained voices and untrained voices. No significant difference between groups was observed in gender discrimination and voice identity recognition tests
Schelinski et al., 2016	HFA n=16 (13 male; AR=20-51) TD n=16 (13 male; AR=18-52)	VP: (1) Unfamiliar voice discrimination test, (2) Unfamiliar voice learning test (3) Famous voice recognition test (4) Acoustic voice features processing test. (5) Musical pitch discrimination test	HFA<TD lower recognition and discrimination of unfamiliar voices, while famous voice recognition was relatively similar. HFA=TD intact timbre discrimination and musical pitch recognition.

Table 3.1: A detailed description of each behavioral studies examining identity information processing in individuals with ASD. In total, poor voice recognition is observed in ASD when linguistic units are large, i.e., sentences, as well as when discriminating unfamiliar voices or familiar voice without a previous training. Legends: GD- Gender Discrimination; VIR – Vocal Identity Recognition; VP – Voice Processing test.

Neuroimaging studies (see Table 3.2) show that when presented with sentences whose content is congruent with the speaker's age, gender or social background, both adults with ASD and typically developing controls significantly activated the left superior temporal gyrus (Tesink et al., 2009). However, when listening to speech information

that is incongruent with speaker's characteristics, different brain regions were activated: right ventromedial prefrontal cortex and right anterior cingulate cortex in controls, but right inferior frontal gyrus in individuals with ASD (Tesink et al., 2009). The ventromedial prefrontal cortex is involved in self-referential processing related to judgments and inferences about the self and others: functional abnormalities in this region are associated with deficits in emotional and social functioning (Kim & Johnson, 2013).

The recruitment of the medial prefrontal cortex is also observed during emotional prosody processing, particularly during the interpretation of complex emotions (Alba-Ferrara et al., 2011). The reduced activation of this brain region in individuals with ASD possibly points to atypical self-referential processing, as well as to abnormal attribution of mental states. The self-other distinction is also important to be made in order to understand another point of view and behavior and to develop social cognition capacities (Malle, 2002). This is crucial during social communication where people share their own perspectives about other people, things or situations, which not always are compatible with our mental states (de Villers & de Villers, 2014).

Difficulties with recognition of identity information may result from the concurrent verbal content, which may interfere with the perceptual discrimination of the speaker identity information. When presented with voice identity information vs. speech stimuli, the activation of the posterior superior temporal sulcus/gyrus is observed in adults with ASD but not in its anterior portion as in typically developing controls (Schelinski, Borowiak, & von Kriegstein, 2016). The reduced response within voice-sensitive areas specialized for the processing of vocal sounds suggests that familiar voices are perceived differently by individuals with ASD.

Publication	Sample	Task	Results
Tesink et al., 2009	HFA n=24 (16 males; AR=18–40) TD n= 24 (16 males; AR=18–39)	Discrimination task from sentences congruent/ incongruent with speaker's age, gender or social background)	HFA=TD significantly activated left STG for speaker congruent sentences. HFA>TD stronger activation in the right IFG for speaker-incongruent sentences vs. speaker-congruent. Only TD showed decreased activation for speaker-incongruent vs. congruent sentences in right vMPFC including right accumbens.

Schelinski et al., 2016	HFA N=16 (13 male; AR=20-51) TD n=16 (13 male; AR=18-52)	Vocal sounds task: Words and vocalizations (laughs and sighs). vs. non-vocal sounds (car sounds, wind, birdsong, saxophone) VIR participants memorised the target speaker and indicated for each sentence whether it was spoken by the same speaker or not	HFA=TD typical responses in the superior temporal sulcus/gyrus (STS/G) for passive listening to vocal sounds vs. non-vocal sounds. (ii) HFA<TD showed less activation of the right posterior STG for voice identity recognition

Table 3.2: A detailed description of each neuroimaging studies examining identity information processing in individuals with ASD. In total, poor voice recognition is observed in ASD when linguistic units are large, i.e., sentences, as well as when discriminating unfamiliar voices or familiar voice without a previous training. Legends: VIR – Voice Identity Recognition; vMPFC – Ventromedial Prefrontal Cortex.

4.4 ASD and Oxytocin

Auditory preference towards human voice is important for social interaction and communication skills development, which serve as salient vocal cues socially relevant events. The mother's voice represents the most attention-grabbing and informative vocal input for infants as compared to other female voices (DeCasper & Fifer, 1980). Listening to one's own mother's vocalizations, particularly in stressful situations, elicits the highest levels of oxytocin release in the typically developing infant's brain. This suggests that preference to listen to the mother's voice is important for the neuroendocrine regulation of social bonding in our species (Seltzer, Ziegler, & Pollak, 2010), which seems to be impaired in individuals with ASD, who show difficulties in developing and maintaining social reciprocity.

Oxytocin is a neuropeptide mainly synthesized by the paraventricular and supraoptic nuclei neurons in the hypothalamus and released into the peripheral circulation after its storage in the posterior lobe of the pituitary gland (Zhang & Han, 2017). From hypothalamus there is a subsequent neuron direct projection to other subcortical brain regions such as the amygdala, striatum, and hippocampus (Meyer-Lindenberg et al., 2011), where amygdala plays a crucial role for emotional information processing. In typically developing individuals, oxytocin modulates different aspects of human social cognition and prosocial behavior, (Kumar, Völlm, & Palaniyappan, 2015; Yatawara et al., 2016). Enhanced rewarding valence of social stimuli, empathy, and willingness to

interact with others (Quattrocki & Friston, 2014) is observed after oxytocin administration, suggesting its strong influence in social interaction skills as well as in emotion and reward processing (Gordon et al., 2016).

The observed fact that children with ASD do not show preference towards their mother's voice (Klin, 1991), may influence the regulation of the oxytocin system in individuals with ASD. On the other hand, the lack of auditory preference towards socially relevant vocal sounds may be explained by atypical genetic and/or epigenetic oxytocin regulation in ASD (Quattrocki & Friston, 2014), who present low plasma endogenous oxytocin levels (Husarova et al., 2016; Modahl et al., 1998). These results indicate oxytocin dysfunction very early in development, with no explanation to how primary dysfunction of this system occurs (Quattrocki et al., 2014). Possibly, the low level of oxytocin release in individuals with ASD prevents the establishment of early parent-child bounding (Seltzer, Ziegler, & Pollak, 2010), and consequently of development of social reciprocity – which seems reflected in social interaction and communication skills deficits present in ASD (American Psychiatric Association, 2013).

The exogenous oxytocin administration serves as a good candidate for alleviating social impairments present in this neurodevelopmental disorder. It improves emotion recognition for less socially proficient individuals, i.e., those who found a task more demanding or challenging, putatively by allocating attentional resources to social stimuli, which turns helpful for individuals with ASD (Bartz et al., 2011; Quattrocki & Friston, 2014). Those studies included in the present systematic review (see Table 4) show that the administration of oxytocin in children and adolescents with ASD before happy voice processing, results in increased connectivity between nucleus accumbens and posteromedial portion of the parietal lobe, known as precuneus, indicating the upregulated connectivity along a major reward pathway after oxytocin administration (Gordon et al., 2016). These brain regions play a role in refining action selection and promoting approach towards motivationally relevant stimuli (Floresco, 2015), as well as during self-relevant information processing, e.g., judgments on one's own versus another person face/personality traits (Cavanna & Trimble, 2006). Thus, a deficit of endogenous oxytocin in ASD may shed light on why happy prosody seems to be the least accurately recognized emotional category by individuals with ASD (Baker,

Montgomery, & Abramson, 2010; Globerson et al., 2015; Grossman et al., 2010; Wang & Tsao, 2015).

The upregulated connectivity between the right anterior insula and superior temporal sulcus is also observed in individuals with ASD while inferring others' social emotions, and in the dorsomedial prefrontal cortex while inferring beliefs (Aokii et al., 2014). Moreover, exogenous oxytocin administration reduces the localized connectivity of the right amygdala, which seems to be enlarged in children with ASD (Schumann et al., 2009a) and which functional connectivity with the bilateral medial prefrontal cortex plays a critical role for social and communication skills development (Shen et al., 2016). Therefore, there is a prosocial effect of exogenous oxytocin administration in ASD by reducing negative reactions and aversive associative learning in response to vocal sounds, for example enhancing access to memories of previous experience with happy cues and their reward value (Gordon et al., 2016; J. Kumar et al., 2015).

Publication	Sample	Task	Results
Gordon et al., 2016	HFA n=20 (17 male; AR=8 - 16.5) No comparison TD group	AV that contrasted listening to a rewarding versus aversive social stimulus (happy versus angry voices.	Intranasal oxytocin increased connectivity between nucleus accumbens and aSMG/HG and PCN during happy but not angry voices in HFA. Perception of happy voices elicited effective connectivity between AMG and PCN
Aokii et al., 2014	HFA n=17 (13 male; AR=24-41) TD n=17 (AR= 20-44)	Sally-Ann false belief task: Participants were required to answer questions about social emotion: "does she feel playful, seeing the box opened? Belief: "does she look for her ball in the box?" and control question: "Is actually the ball in the box?"	HFA<TD lower activation of the right anterior insula and STS while inferring others' social emotions, and in the dMPFC while inferring beliefs. Oxytocin administration enhanced activity in the right anterior insula in individuals with HFA.

Table 4: From a total of 41 only 2 studies have examined the effects of exogenous oxytocin administration over vocal information processing in individuals with ASD. In total, intranasal oxytocin administration in individuals with ASD seems to improve functional connectivity between cortical and subcortical brain regions involved during emotional stimuli processing, e.g., amygdala, insula nucleus accumbens. Legends: AV – Affective Voices test; aSMG – anterior Supramarginal Gyrus; HG – Heschl's Gyrus; PCN – Precuneus; AMG – Amygdala; dMPFC – dorsomedial Prefrontal Cortex.

5. DISCUSSION

Since the early work of Leo Kanner and his publication on *Autistic Disturbances of Affective Contact* (1943), the reported children with Autism showed disturbances when perceiving *loud noises or motions that intrude itself upon the child aloneness* (Kanner, 1943, p.. 245). The absence of automatic orienting towards socially relevant sensory vocal stimuli, is one of the key signs of abnormal perception of social cues from the voice in individuals with ASD, which is reflected on social interaction and communication deficits (American Psychiatric Association, 2013).

Unlike typically developing infants who show sensitivity to prosodic characteristics of speech (Gervain & Mehler, 2010), who attend selectively to speech sounds over many other naturally occurring auditory stimuli (Kuhl, et al., 2005; Shultz & Vouloumanos, 2010), who recognize their own names and isolated words in the speech stream (Brooks & Kempe, 2012), children with ASD do not show an automatic bias for socially relevant vocal stimuli (Magrelli et al., 2013). On the contrary, the auditory preference in individuals with ASD is usually observed over non-speech stimuli (Kuhl et al., 2005; Vouloumanos & Werker, 2007), showing no automatic attentional orienting to emotional vocal stimuli (Blasi et al., 2015) or to the mother's voice (Klin, 1991), exhibiting difficulties in extracting mental states information from the tone of voice (Shultz et al., 2012).

As mentioned earlier in section 1.1.1, both mother's voice and infant-directed speech are characterized by high frequency modulations that seem to be advantageous for typically developing infants whose frequency discrimination abilities are better in the region of 500 Hz as compared to 100 Hz (Fernald, 1985). However, the reduced auditory preference to infant-directed speech in pre-school children with ASD and infants at risk for developing ASD, who spend more time listening to distorted and non-speech analogue signals instead (Kuhl et al., 2005; Paul, et al., 2007), suggests that the auditory system may not be attracted by higher pitch contours, which may interfere with further processing of vocal affective stimuli, for example.

A recent neuroimaging study of Abrams and collaborators (2016) indicates that when listening to the mother's voice, a functional connection between voice-selective brain regions and the reward system, including the amygdala, is observed in the typically

developing infant's brain (Abrams et al., 2016). However, reduced functional connectivity between the left posterior superior temporal gyrus and nucleus accumbens, as well as between the right posterior superior temporal gyrus and amygdala together with orbitofrontal cortex is also observed in ASD (Abrams et al., 2013), suggesting the abnormal interhemispheric functioning of the posterior portion of the superior temporal gyrus in ASD. As previously mentioned, the posterior portion of voice sensitive areas constitutes one of the key brain regions for the initial analysis of the phonemic level information especially in the left hemisphere (Buckingham H. W. et al., 2001). Thus, abnormal early auditory processing of vocal sounds may interfere with subsequent integration steps when sensory low-level analysis is combined with the semantic meaning or emotional value in high-order information processing stages.

Is voice information processing disrupted in ASD individuals?

Considering the model of voice perception proposed by Belin and collaborators (2004), it seems that deficits with voice information processing in individuals with ASD arise at a bottom-up perceptual stage of voice perception, in which the general low-level auditory analysis takes place (see Figure 3). The initial low-level acoustic analysis seems to be degraded in this specific population, interfering further in the structural encoding of the vocal stimuli. This hypothesis is corroborated by neuroimaging studies showing the reduced activity within temporal voice-sensitive areas in individuals with ASD (Blasi et al., 2015; Gervais et al., 2004; Schelinski et al., 2016). The abnormal functioning and the early absence of cortical specialization for voice information processing in infants at risk for developing autism (Blasi et al., 2015), suggests a different neural processing of voice information by individuals with ASD compared to typically developing controls. The abnormal functioning of voice-sensitive areas in individuals with ASD may serve as the main cause for the reduced auditory preference for human vocal sounds in ASD (Blasi et al., 2015; Gervais et al., 2004; Schelinski et al., 2016).

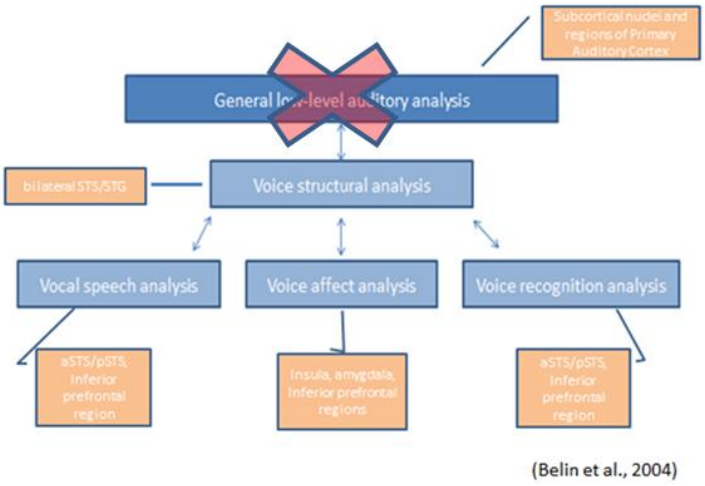

Figure 3: The early impairment during general acoustic analysis of the vocal sounds in individuals with ASD seems to prevent forward evaluation of speech, affect and identity information from the voice.

A recent animal study of Anomal and collaborators (2015) indicates abnormal primary auditory cortex functioning in a rat model of autism, whose tonotopic organization seems to be distorted with over-representation of high frequencies (Anomal et al., 2015). As described earlier in section 1.1.3, the primary auditory cortex, and more specifically the supra-temporal plane, plays a key role during the initial analysis of spectro-temporal acoustic parameters of complex auditory stimuli e.g., speech, (Zatorre & Belin, 2001), which functioning is crucial for the perception of formant frequencies (Formisano et al., 2008) important not only for phoneme information processing, but also for identity information. The recruitment of the right supra-temporal plane, in particular, occurs during female voice processing together with the insula (Lattner et al., 2005), whose dysfunction may result in reduced auditory preference for high pitched vocal sounds.

Is speech information processing abnormal in ASD individuals?

Relative to speech information, the early lack of tuning to and discrimination of native speech sounds in children with ASD (Ceponiené et al., 2003; DePape, et al., 2012; Gervais et al., 2004; Lepistö et al., 2005; Whitehouse & Bishop, 2008, Seery et al., 2013) point to abnormalities in phonological information processing, which may influence language receptive and expressive abilities. This difficulty may be a result of the reduced cortical activity in the left hemisphere during speech processing observed in

individuals with ASD, who show a clear rightward asymmetry (Eyler, Pierce, & Courchesne, 2012; Floris et al., 2016).

The predominant right hemisphere involvement may result as a spare mechanism for segmental information processing, because of the volumetric enlargement of the right superior temporal gyrus, and more specifically of its posterior portion observed in individuals with ASD (Jou et al., 2010). An abnormal functioning of language-related regions in individuals with ASD may result from the reduced left superior longitudinal fasciculus as suggested by neuroimaging evidence (Sharda et al., 2014). The disrupted activity within this particular brain area may influence the subsequent processing of social information within the anterior superior temporal cortex, as the core center for social cognition tasks, including the theory of mind or mentalizing (Olson et al., 2013).

Is there a neural pathway for voice information processing specifically impaired in individuals with Autism?

After the analysis of findings coming from behavioral, neuroimaging and electrophysiological studies, it seems that the brain pathway for vocal affect information processing is the most impaired in ASD population. Specific deficits during the decoding of vocal affect information in individuals with ASD are observed in the case of happy vocal sounds (Baker et al., 2010; Brennand, Schepman, & Rodway, 2011; Le Sourn-Bissaoui, et al., 2013; Wang & Tsao, 2015). The infant's attentional preference towards adult-adult conversation is often attracted by the positive affect that it contains (Saint-Georges et al., 2013). Therefore, the auditory orienting towards high pitched voices in typically developing children suggests its relevance as socially orienting cues. However, the reduced auditory preference towards the mother's voice in children with ASD and poorer recognition of happy valence, suggest the abnormal processing of higher frequencies in individuals with ASD characterized by fast fundamental frequency modulations (Lattner et al., 2005). Moreover, whereas in typically developing children, the processing of high-frequency tones is observed in the left hemisphere (Demopolous et al., 2015), the predominant activation of the right hemisphere in individuals with ASD corroborates the abnormal processing of this acoustic parameter by ASD population.

Difficulties in recognizing emotional vocal stimuli characterized by higher pitched tone of voice may result from abnormalities of the right amygdala, which is enlarged in individuals with ASD (Nordahl et al., 2012; Schumann et al., 2009; Shen et al., 2016). Interestingly, a longitudinal study of Ortiz-Mantilla and collaborators (2010) showed that those children who present larger volume of the right amygdala had lower outcomes in receptive and expressive language skills forward on developmental trajectory (Ortiz-Mantilla et al., 2010). It has been suggested that the amygdala helps the individual to detect socio-emotionally salient stimuli, it heightens arousal in response to these stimuli, and it facilitates learning and remembering their reward value (Gordon et al., 2016). Moreover, it is functionally connected with the orbitofrontal cortex (Abrams et al., 2013; Fossati, 2012), which plays an important role not only during vocal affect information processing (Schirmer & Kotz, 2006) but is also important in theory of mind functions which seem to be disrupted in individuals with ASD (Saxe, 2009). The observed underconnectivity between these brain regions in individuals with ASD (Abrams et al., 2013) suggest the abnormal processing of self-relevant vocal information in ASD (Fruhholz, Trost, & Kotz, 2016; Leitman et al., 2016).

Because valence modulates the backward connections from the amygdala to the auditory cortex (Kumar et al., 2012), the lack of motivation towards positive valence in individuals with ASD may interfere with the functioning of voice-sensitive areas, resulting in lower emotional recognition in individuals with ASD. Altogether, there are two possible brain regions that might be responsible for the abnormal processing of vocal affect information in individuals with ASD: the superior temporal gyrus/sulcus and the right amygdala. We still do not have robust evidence about which of these brain structures interfere with the typical functioning of the other; is it the amygdala's abnormal functioning that prevents the typical activation of voice-sensitive regions, or is it the temporal brain areas impairments that interfere with the evaluation and attribution of reward value to emotional vocal sounds (see Figure 4). Although there is neuroimaging evidence supporting an abnormally reduced activation of critical temporal areas for voice processing in individuals, it is still important to investigate the possible modulation of subcortical areas over the functioning of voice-sensitive brain regions.

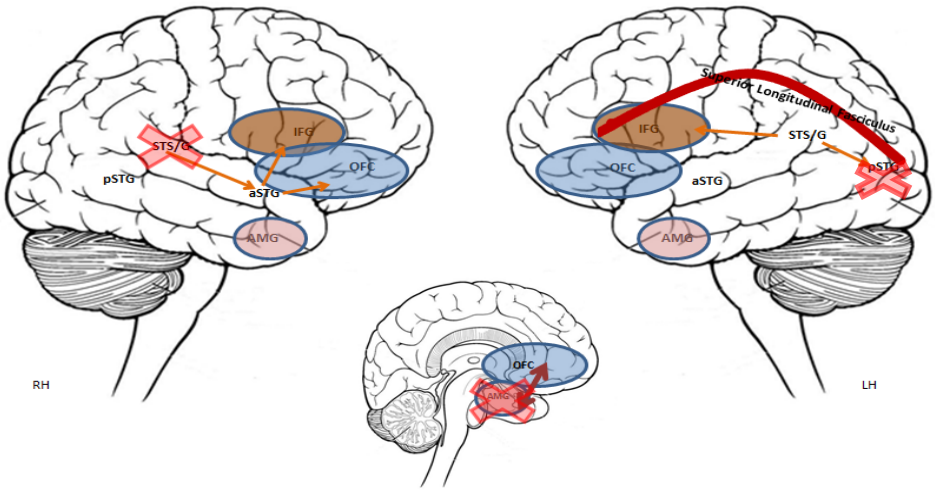

Figure 4: The present figures serves as explanatory hypotheses of the disrupted functioning of possible cortical and subcortical areas in ASD. When perceiving a voice, an initial low-level analysis of the voice in the Superior Temporal Gyrus/Sulcus (STS/G) may be disrupted, with further consequences for the attribution of emotional value in prefrontal areas predominantly in the right hemisphere. Because both cortical and subcortical areas are activated during emotional information processing, the disrupted functional connection between the amygdala and orbitofrontal cortex may interfere with the evaluation of emotional value. On the other hand, the abnormal voice-sensitive areas may interfere with the attribution of emotional meaning forward within the inferior frontal cortex. When processing speech information, the initial analysis of fast phonemic information within the posterior portion of the STG may be impaired due to the lack of tuning to the native sounds of speech. Moreover, the abnormal functioning of the posterior areas of left hemisphere may result from a reduced voxel-wise fractional anisotropy of the left superior longitudinal fasciculus, which serves as a crucial neurobiological pathway connecting language-relevant brain regions. In sum, more scientific research is needed in order to understand which brain structure contributes for the abnormal processing of vocal information processing in ASD.

Is there a possible improvement effect via exogenous oxytocin administration?

The abnormal functioning and connectivity along major reward pathway with cortical brain regions may be complemented with the neurobiological evidence of abnormal endogenous oxytocin system in individuals with ASD. Given that oxytocin is crucial for parent-infant bonding (Quattrocki & Friston, 2014), its low levels in ASD may serve as an additional explanatory hypothesis of the reduced preference for listening to vocal sounds characterized by high pitch such as female/mother's voice, leading to abnormal orienting towards socially relevant information. Some of the studies suggest

improvement in emotional tone of voice decoding in individuals with ASD when the exogenous oxytocin is administered (Aoki et al., 2014; Gordon et al., 2016). However, single-administration studies demonstrate only short-term symptom improvement in individuals with ASD; whereas chronic administration studies are scarce and show mixed effects (Gordon et al., 2016). Although there is already a systematic review of Quattrocki and Friston (2014), as well as of Ooi and collaborators (2016) about oxytocin exogenous administration in ASD population, it is still important to clarify how oxytocin may improve deficits in social cognition and emotional vocal stimuli processing present in this neurodevelopmental disorder. More clinical proof is needed to look at whether endogenous oxytocin levels moderate the response to exogenously administered oxytocin in individuals with ASD, with possible improvement in the decoding of emotional prosody and social cognition skills in ASD.

Is there a possible relationship of abnormal vocal affect information processing and impaired Theory of Mind in ASD?

The correct decoding of emotional tone of voice is crucial to accurately infer an individual's dispositions and intentions (Zilbovicius et al., 2013) and explain one's actions (Bzdok et al., 2016), as well as to understand the linguistic non-linear meaning, e.g., irony, sarcasm (Speer & Ito, 2009). This cognitive ability coined as 'theory of mind' by Premack and Woodruff in 1978 in the study of *Does the chimpanzee have a theory of mind?*, has been progressively replicated in children in order to understand how they attribute mental states (desires, beliefs etc.) to themselves and others, by distinguishing intentions from accidents based on socio-pragmatic cues, such as prosody information (Sakkalou & Gattis, 2012).

At the brain level, the socio-cognitive skills including mentalizing, reasoning and inference about other's states of mind require the activation of specific brain regions, such as the right temporo-parietal junction and medial prefrontal cortex (Saxe, 2009). Nevertheless, individuals with ASD have difficulty with attribution and comprehension of mental states (Baron-Cohen, 2001; O'Nions et al., 2014), a process which requires adjudicating an internal cognitive state as well as an emotion to a social context (Alba-Ferrara et al., 2011). The disrupted theory of mind in ASD may result from the reduced activity of the right temporoparietal junction observed in individuals with ASD during

mentalizing about self and others (Lombardo et al., 2011). More specifically, local connectivity within the posterior right temporoparietal junction seems to be disrupted in children with ASD, which is functionally connected to the dorsomedial prefrontal subsystem crucial for self-relevant social information processing (Chien et al., 2015) .

On the other hand, subcortical brain regions such as the basal ganglia have also been suggested to be important for inferential processing of recognizing the feelings of another person, which associated disorders have been reported to elicit theory of mind deficits (Bodden et al., 2013). The activation of basal ganglia is observed during affective sounds processing, indicating its role in emotional prosody decoding (Fruhholz, Trost, & Kotz, 2016; Schirmer & Kotz, 2006). Moreover, basal ganglia receptive neurons are sensitive to oxytocin neuropeptide release, which mediates social interaction and prosocial behavior, suggesting its important role both in oxytocin regulation and Theory of Mind development.

In sum, the abnormal processing of emotional tone of voice in ASD may result in impaired processing o complex social emotions, which decoding depends on accurate perception of prosodic information, essential during social interaction and communication.

6. CONCLUSION

The present systematic review presented interdisciplinary evidence about abnormal voice and speech information processing in ASD. After a careful analysis of each scientific outcome, the impaired processing of vocal affect information in particular was established. It seems that a higher pitched tone of voice, such as a female voice (e.g., the mother's voice) and positive emotional valence do not serve as a saliently enough cue to guide an infant towards socially relevant information in the environment. Because infant's perception of pitch deteriorates as the number of harmonics in a tonal complex stimuli decreases (Clarkson, Martin, & Miciek, 1996), the lower number of harmonics present in high frequency vocal stimuli may be difficult to be perceived by children with ASD, whose auditory system seems to be disrupted when processing acoustic parameters such as fundamental frequency.

The decoding of suprasegmental information is crucial for successful social interaction and communication skills development from which socially relevant information should be extracted, such as the speaker's mental states. The observable lack of preference towards vocal sounds may have its origin in the absence of early cortical specialization for voice processing in temporal voice-sensitive areas and its abnormal connectivity with reward, affective and salience regions. On the other hand, disrupted release of endogenous oxytocin may interfere with the processing of socially relevant vocal information and prosocial behavior.

In total, more studies are needed in order to obtain more homogeneous evidence about how ASD individuals perceive such a socially relevant stimulus, i.e. the voice, especially during identity information processing. Furthermore, more scientific research is needed investigating the effect of oxytocin during vocal information processing in ASD, due to the lack of a clear improvement effect of exogenous oxytocin administration in individuals with ASD. It is crucial to understand what happens during the developmental trajectory of auditory processing in this Autism Spectrum Disorders, so more accurate and earlier intervention can be made for better social communication and interaction scores achievement.

REFERENCES

Abrams, D. A., Chen, T., Odriozola, P., Cheng, K. M., Baker, A. E., Padmanabhan, A., … Menon, V. (2016). Neural circuits underlying mother's voice perception predict social communication abilities in children. *Proceedings of the National Academy of Sciences*, *113*(22), 6295–6300. https://doi.org/10.1073/pnas.1602948113

Abrams, D. a, Lynch, C. J., Cheng, K. M., Phillips, J., Supekar, K., Ryali, S., … Menon, V. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(29), 12060–5. https://doi.org/10.1073/pnas.1302982110

Alba-Ferrara, L., Hausmann, M., Mitchell, R. L., & Weis, S. (2011). The neural correlates of emotional prosody comprehension: Disentangling simple from complex emotion. *PLoS ONE*, *6*(12). https://doi.org/10.1371/journal.pone.0028701

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. *Arlington*. https://doi.org/10.1176/appi.books.9780890425596.744053

Anne M. Seery, Vogel-Farley, V., Helen Tager-Flusberg, A., & Nelson2, C. A. (2014). NIH Public Access. *Dev Cogn Neurosci*, (617), 10–24. https://doi.org/10.1016/j.dcn.2012.11.007.Atypical

Anomal, R. F., Villers-sidani, E. De, Brandão, J. A., Diniz, R., Costa, M. R., Romcy-pereira, R. N., & Costa, M. R. (2015). Impaired Processing in the Primary Auditory Cortex of an Animal Model of Autism, *9*(November), 1–12. https://doi.org/10.3389/fnsys.2015.00158

Aoki, Y., Yahata, N., Watanabe, T., Takano, Y., Kawakubo, Y., Kuwabara, H., … Yamasue, H. (2014). Oxytocin improves behavioural and neural deficits in inferring others' social emotions in autism. *Brain*, *137*(11), 3073–3086. https://doi.org/10.1093/brain/awu231

Ardila, A., & Byron Bernal. (2016). From Hearing Sounds to Recognizing Phonemes: Primary Auditory Cortex is A Truly Perceptual Language Area. *{AIMS} Neuroscience*, *3*(4), 454--473. https://doi.org/10.3934/Neuroscience.2016.4.454

Baker, K. F., Montgomery, A. A., & Abramson, R. (2010). Brief report: Perception and lateralization of spoken emotion by youths with high-functioning forms of autism. *Journal of Autism and Developmental Disorders*, *40*(1), 123–129. https://doi.org/10.1007/s10803-009-0841-1

Baron-Cohen, S. (2001). Theory of Mind in Normal Development and Autism. *Prisme*, *34*, 174–183. Retrieved from http://www.autism-community.com/wp-content/uploads/2010/11/TOM-in-TD-and-ASD.pdf

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: Context and person matter. *Trends in Cognitive Sciences*, *15*(7), 301–309. https://doi.org/10.1016/j.tics.2011.05.002

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of

voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–135. https://doi.org/10.1016/j.tics.2004.01.008

Belin, P., & Grosbras, M. H. (2010). Before Speech: Cerebral Voice Processing in Infants. *Neuron*, *65*(6), 733–735. https://doi.org/10.1016/j.neuron.2010.03.018

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26. https://doi.org/10.1016/S0926-6410(01)00084-2

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312. https://doi.org/10.1038/35002078

Bethmann, A., Scheich, H., & Brechmann, A. (2012). The Temporal Lobes Differentiate between the Voices of Famous and Unknown People: An Event-Related fMRI Study on Speaker Recognition. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0047626

Bidet-Caulet, A., Latinus, M., Roux, S., Malvy, J., Bonnet-Brilhault, F., & Bruneau, N. (2017). Atypical sound discrimination in children with ASD as indicated by cortical ERPs. *Journal of Neurodevelopmental Disorders*, *9*(1), 13. https://doi.org/10.1186/s11689-017-9194-9

Blasi, A., Lloyd-Fox, S., Sethna, V., Brammer, M. J., Mercure, E., Murray, L., … Johnson, M. H. (2015). Atypical processing of voice sounds in infants at risk for autism spectrum disorder. *Cortex*, *71*, 122–133. https://doi.org/10.1016/j.cortex.2015.06.015

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., … Murphy, D. G. M. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology*, *21*(14), 1220–1224. https://doi.org/10.1016/j.cub.2011.06.009

Boddaert, N., Chabane, N., Belin, P., Bourgeois, M., Royer, V., Barthelemy, C., … Zilbovicius, M. (2004). Perception of complex sounds in autism: Abnormal auditory cortical processing in children. *American Journal of Psychiatry*, *161*(11), 2117–2120. https://doi.org/10.1176/appi.ajp.161.11.2117

Bodden, M. E., Kübler, D., Knake, S., Menzler, K., Heverhagen, J. T., Sommer, J., … Dodel, R. (2013). Comparing the neural correlates of affective and cognitive theory of mind using fMRI: Involvement of the basal ganglia in affective theory of mind. *Advances in Cognitive Psychology*, *9*(1), 32–43. https://doi.org/10.2478/v10053-008-0129-6

Brennand, R., Schepman, A., & Rodway, P. (2011). Vocal emotion perception in pseudo-sentences by secondary-school children with Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *5*(4), 1567–1573. https://doi.org/10.1016/j.rasd.2011.03.002

Buckingham H. W., J., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, *25*(5), 663–678. https://doi.org/10.1016/S0364-

0213(01)00048-9

Bzdok, D., Hartwigsen, G., Reid, A., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2016). Left inferior parietal lobe engagement in social cognition and language. *Neuroscience and Biobehavioral Reviews*, *68*, 319–334. https://doi.org/10.1016/j.neubiorev.2016.02.024

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, *129*(3), 564–583. https://doi.org/10.1093/brain/awl004

Ceponiené, Lepisto, Shestakiva, Vanhala, Alku, Naatanen, & Yaguchi. (2003). Speech – sound-selective auditory impairment in children with autism : They can perceive but ˇ. *Pnas*, *100*(9), 5567–5572.

Chien, H. Y., Lin, H. Y., Lai, M. C., Gau, S. S. F., & Tseng, W. Y. I. (2015). Hyperconnectivity of the Right Posterior Temporo-parietal Junction Predicts Social Difficulties in Boys with Autism Spectrum Disorder. *Autism Research*, *8*(4), 427–441. https://doi.org/10.1002/aur.1457

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*, *31*(3–4), 585–598. https://doi.org/10.1016/S0095-4470(03)00040-8

Clarkson, M. G., Martin, R. L., & Miciek, S. G. (1996). Infants ' Perception of Pitch : Number of Harmonics, 191–197.

Cutler, A. (2012). Native listening: The flexibility dimension. *Dutch Journal of Applied Linguistics*, *1*(2), 169–187. https://doi.org/10.1075/dujal.1.2.02cut

de Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal Attention as a Scaffold for Language Development. *Frontiers in Psychology*, *7*(February), 1–15. https://doi.org/10.3389/fpsyg.2016.00044

de Villiers, J. G., & de Villiers, P. A. (2014). The Role of Language in Theory of Mind Development. *Topics in Language Disorders*, *34*(4), 313–328. https://doi.org/10.1097/TLD.0000000000000037

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*. https://doi.org/10.1126/science.7375928

Dehaene-lambertz, G., Dehaene-lambertz, G., & Dehaene, S. (2014). Functional Neuroimaging of Speech Perception in Infants, *2013*(2002). https://doi.org/10.1126/science.1077066

Demopoulos, C., Ph, D., Hopkins, J., Ph, D., Kopald, B. E., Psy, D., … Ph, D. (2016). Study, *29*(6), 895–908. https://doi.org/10.1037/neu0000209.Deficits

DePape, A.-M. R., Hall, G. B. C., Tillmann, B., & Trainor, L. J. (2012). Auditory Processing in High-Functioning Adolescents with Autism Spectrum Disorder. *PLoS ONE*, *7*(9), e44084. https://doi.org/10.1371/journal.pone.0044084

Dunlop, W. A., Enticott, P. G., & Rajan, R. (2016). Speech Discrimination Difficulties in High-Functioning Autism Spectrum Disorder Are Likely Independent of

Auditory Hypersensitivity. *Frontiers in Human Neuroscience*, *10*(August), 1–12. https://doi.org/10.3389/fnhum.2016.00401

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex*, *22*(1), 191–200. https://doi.org/10.1093/cercor/bhr113

Eyler, L. T., Pierce, K., & Courchesne, E. (2012). A failure of left temporal cortex to specialize for language is an early emerging and fundamental property of autism. *Brain*, *135*(3), 949–960. https://doi.org/10.1093/brain/awr364

Fan, Y. T., & Cheng, Y. (2014). Atypical mismatch negativity in response to emotional voices in people with autism spectrum conditions. *PLoS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0102471

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Floresco, S. B. (2015). The Nucleus Accumbens: An Interface Between Cognition, Emotion, and Action. *Annual Review of Psychology*, *66*(1), 25–52. https://doi.org/10.1146/annurev-psych-010213-115159

Floris, D. L., Lai, M., Auer, T., Lombardo, M. V, Ecker, C., Chakrabarti, B., … Murphy, D. G. M. (2016). Atypically Rightward Cerebral Asymmetry in Male Adults With Autism Stratifies Individuals With and Without Language Delay, *253*, 230–253. https://doi.org/10.1002/hbm.23023

Fossati, P. (2012). Neural correlates of emotion processing: From emotional to social brain. *European Neuropsychopharmacology*, *22*(SUPPL3), S487–S491. https://doi.org/10.1016/j.euroneuro.2012.07.008

Fr??hholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions-Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews*, *68*, 1–15. https://doi.org/10.1016/j.neubiorev.2016.05.002

Fridenson-Hayo, S., Berggren, S., Lassalle, A., Tal, S., Pigat, D., Bölte, S., … Golan, O. (2016). Basic and complex emotion recognition in children with autism: cross-cultural findings. *Molecular Autism*, *7*(1), 52. https://doi.org/10.1186/s13229-016-0113-9

Friederici, A. D. (2012). The cortical language circuit : from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16*(5), 262–268. https://doi.org/10.1016/j.tics.2012.04.001

Fujii, T., Fukatsu, R., Watabe, S. ichi, Ohnuma, A., Teramura, K., Kimura, I., … Kogure, K. (1990). Auditory Sound Agnosia without Aphasia Following a Right Temporal Lobe Lesion. *Cortex*, *26*(2), 263–268. https://doi.org/10.1016/S0010-9452(13)80355-3

Gebauer, L., Skewes, J., Hørlyck, L., & Vuust, P. (2014). Atypical perception of affective prosody in Autism Spectrum Disorder. *NeuroImage: Clinical*, *6*, 370–

378. https://doi.org/10.1016/j.nicl.2014.08.025

Gervain, J., & Mehler, J. (2010). Speech Perception and Language Acquisition in the First Year of Life. *Annual Review of Psychology*, *61*(1), 191–218. https://doi.org/10.1146/annurev.psych.093008.100408

Gervain, J., & Werker, J. F. (2008). Infant Speech Perception Contributes to Language Acquisition, *6*, 1149–1170. https://doi.org/10.1111/j.1749-818X.2008.00089.x

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., … Zilbovicius, M. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience*, *7*(8), 801–802. https://doi.org/10.1038/nn1291

Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015). Prosody recognition in adults with high-functioning autism spectrum disorders: From psychoacoustics to cognition. *Autism Research*, *8*(2), 153–163. https://doi.org/10.1002/aur.1432

Goense, J., Bohraus, Y., & Logothetis, N. K. (2016). fMRI at High Spatial Resolution: Implications for BOLD-Models. *Frontiers in Computational Neuroscience*, *10*(June), 1–13. https://doi.org/10.3389/fncom.2016.00066

Gordon, I., Jack, A., Pretzsch, C. M., Vander Wyk, B., Leckman, J. F., Feldman, R., & Pelphrey, K. A. (2016). Intranasal Oxytocin Enhances Connectivity in the Neural Circuitry Supporting Social Motivation and Social Perception in Children with Autism. *Scientific Reports*, *6*(1), 35054. https://doi.org/10.1038/srep35054

Groen, W. B., Van Orsouw, L., Huurne, N. Ter, Swinkels, S., Van Der Gaag, R. J., Buitelaar, J. K., & Zwiers, M. P. (2009). Intact spectral but abnormal temporal processing of auditory stimuli in autism. *Journal of Autism and Developmental Disorders*, *39*(5), 742–750. https://doi.org/10.1007/s10803-008-0682-3

Grossman, R. B., Bemis, R. H., Plesa Skwerer, D., & Tager-Flusberg, H. (2010). Lexical and Affective Prosody in Children With High-Functioning Autism. *Journal of Speech Language and Hearing Research*, *53*(3), 778. https://doi.org/10.1044/1092-4388(2009/08-0127)

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The Developmental Origins of Voice Processing in the Human Brain. *Neuron*, *65*(6), 852–858. https://doi.org/10.1016/j.neuron.2010.03.001

Hickok, G., & Poeppel, D. (2007). Processing, *8*(May), 393–402. https://doi.org/10.1038/nrn2113

Homae, F., Watanabe, H., Nakano, T., & Taga, G. (2007). Prosodic processing in the developing brain. *Neuroscience Research*, *59*(1), 29–39. https://doi.org/10.1016/j.neures.2007.05.005

Hruby, T., & Marsalek, P. (2003). Event-Related Potentials - the P3 Wave, 55–63.

Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., & Chang, E. F. (2016). Human Superior Temporal Gyrus Organization of Spectrotemporal Modulation Tuning Derived from Speech Stimuli. *Journal of Neuroscience*, *36*(6), 2014–2026. https://doi.org/10.1523/JNEUROSCI.1779-15.2016

Husarova, V. M., Lakatosova, S., Pivovarciova, A., Babinska, K., Bakos, J., Durdiakova, J., … Ostatnikova, D. (2016). Plasma oxytocin in children with autism and its correlations with behavioral parameters in children and parents. *Psychiatry Investigation*, *13*(2), 174–183. https://doi.org/10.4306/pi.2016.13.2.174

Jacobsen, T., Schröger, E., & Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology*, *41*(4), 654–659. https://doi.org/10.1111/1469-8986.2004.00175.x

Järvinen-Pasley, A., Pasley, J., & Heaton, P. (2008). Is the linguistic content of speech less salient than its perceptual features in autism? *Journal of Autism and Developmental Disorders*, *38*(2), 239–248. https://doi.org/10.1007/s10803-007-0386-0

Jou, R. J., Minshew, N. J., Keshavan, M. S., Vitale, M. P., & Hardan, A. Y. (2010). Enlarged right superior temporal gyrus in children and adolescents with autism. *Brain Research*, *1360*, 205–212. https://doi.org/10.1016/j.brainres.2010.09.005

Jusczyk, P. W. (1999). How infants begin to extract words from speech, *3*(9), 323–328.

Kargas, N., López, B., Reddy, V., & Morris, P. (2015). The Relationship Between Auditory Processing and Restricted, Repetitive Behaviors in Adults with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *45*(3), 658–668. https://doi.org/10.1007/s10803-014-2219-2

Kim, K., & Johnson, M. K. (2013). Activity in ventromedial prefrontal cortex during self-related processing: Positive subjective value or personal significance? *Social Cognitive and Affective Neuroscience*, *10*(4), 494–500. https://doi.org/10.1093/scan/nsu078

Klin, A. (1991). Young Autistic Childrens Listening Preferences in Regard to Speech : A Possible Characterization of the Symptom of Social Withdrawal 1, *21*(1).

Kotz, S. A., Kalberlah, C., Friederici, A. D., & Haynes, J. (2012). Predicting Vocal Emotion Expressions from the Human Brain, *0*(September 2011). https://doi.org/10.1002/hbm.22041

Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, *22*(2), 948–955. https://doi.org/10.1016/j.neuroimage.2004.02.020

Kuhl, P. K. (2004). Early Language Acquistion :, *5*(November). https://doi.org/10.1038/nrn1533

Kuhl, P. K., Kuhl, P. K., Coffey-corina, S., Coffey-corina, S., Padden, D., Padden, D., … Dawson, G. (2005). Links between social and linguistic processing of speech in children with autism: behavioural and electrophysiological measures. *Developmental Science*, *8*(1), 1–12. https://doi.org/10.1111/j.1467-7687.2004.00384.x

Kumar, J., Völlm, B., & Palaniyappan, L. (2015). Oxytocin affects the connectivity of the precuneus and the Amygdala: A randomized, double-blinded, placebo-controlled neuroimaging trial. *International Journal of Neuropsychopharmacology*, *18*(5), 1–7. https://doi.org/10.1093/ijnp/pyu051

Kumar, S., von Kriegstein, K., Friston, K., & Griffiths, T. D. (2012). Features versus Feelings: Dissociable Representations of the Acoustic Features and Valence of Aversive Sounds. *Journal of Neuroscience*, *32*(41), 14184–14192. https://doi.org/10.1523/JNEUROSCI.1759-12.2012

Lancker, V., Lancker, V., & Tartter, V. C. (1987). Acoustic parameters in human speaker recognition", *33*(3), 259–272.

Lartseva, A., Dijkstra, T., Kan, C. C., & Buitelaar, J. K. (2014). Processing of emotion words by patients with autism spectrum disorders: Evidence from reaction times and EEG. *Journal of Autism and Developmental Disorders*, *44*(11), 2882–2894. https://doi.org/10.1007/s10803-014-2149-z

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping*, *24*(1), 11–20. https://doi.org/10.1002/hbm.20065

Le Sourn-Bissaoui, S., Aguert, M., Girard, P., Chevreuil, C., & Laval, V. (2013). Emotional speech comprehension in children and adolescents with autism spectrum disorders. *Journal of Communication Disorders*, *46*(4), 309–320. https://doi.org/10.1016/j.jcomdis.2013.03.002

Leitman, D. I., Edgar, C., Berman, J., Gamez, K., Fruhholz, S., & Roberts, T. P. L. (2016). Amygdala and insula contributions to dorsal-ventral pathway integration in the prosodic neural network, (215), 1–12. Retrieved from http://arxiv.org/abs/1611.01643

Lepistö, T., Kajander, M., Vanhala, R., Alku, P., Huotilainen, M., Näätänen, R., & Kujala, T. (2008). The perception of invariant speech features in children with autism. *Biological Psychology*, *77*(1), 25–31. https://doi.org/10.1016/j.biopsycho.2007.08.010

Lepistö, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Näätänen, R. (2005). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research*, *1066*(1–2), 147–157. https://doi.org/10.1016/j.brainres.2005.10.052

Liebenthal, E., Silbersweig, D. A., & Stern, E. (2016). The language, tone and prosody of emotions: Neural substrates and dynamics of spoken-word emotion perception. *Frontiers in Neuroscience*, *10*(NOV), 1–13. https://doi.org/10.3389/fnins.2016.00506

Lin, I. F., Yamada, T., Komine, Y., Kato, N., Kato, M., & Kashino, M. (2015). Vocal identity recognition in autism spectrum disorder. *PLoS ONE*, *10*(6). https://doi.org/10.1371/journal.pone.0129451

Lindström, R., Lepistö-Paisley, T., Vanhala, R., Alén, R., & Kujala, T. (2016). Impaired neural discrimination of emotional speech prosody in children with autism spectrum disorder and language impairment. *Neuroscience Letters*, *628*, 47–51. https://doi.org/10.1016/j.neulet.2016.06.016

Litovsky, R., & Hearing, B. (2015). *Development of the auditory system*. https://doi.org/10.1016/B978-0-444-62630-1.00003-2.Development

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., & Baron-Cohen, S. (2011). Specialization of right temporo-parietal junction for mentalizing and its relation to social impairments in autism. *NeuroImage*, *56*(3), 1832–1838. https://doi.org/10.1016/j.neuroimage.2011.02.067

Ludlow, A., Mohr, B., Whitmore, A., Garagnani, M., Pulvermüller, F., & Gutierrez, R. (2014). Auditory processing and sensory behaviours in children with autism spectrum disorders as revealed by mismatch negativity. *Brain and Cognition*, *86*(1), 55–63. https://doi.org/10.1016/j.bandc.2014.01.016

Lukatela, G., & Turvey, M. T. (1991). Phonological access of the lexicon: evidence from associative priming with pseudohomophones. *Journal of Experimental Psychology. Human Perception and Performance*, *17*(4), 951–66. https://doi.org/10.1037/0096-1523.17.4.951

Magrelli, S., Jermann, P., Noris, B., Ansermet, F., Hentsch, F., Nadel, J., & Billard, A. (2013). Social orienting of children with autism to facial expressions and speech: A study with a wearable eye-tracker in naturalistic settings. *Frontiers in Psychology*, *4*(NOV), 1–16. https://doi.org/10.3389/fpsyg.2013.00840

Malle, B. F. (2002). The relation between language and theory of mind in development and evolution. *The Evolution of Language out of Prelanguage*, (May 2001), 265–284. Retrieved from http://cogprints.org/3317/

Mayer, J. L., Hannent, I., & Heaton, P. F. (2016). Mapping the Developmental Trajectory and Correlates of Enhanced Pitch Perception on Speech Processing in Adults with ASD. *Journal of Autism and Developmental Disorders*, *46*(5), 1562–1573. https://doi.org/10.1007/s10803-014-2207-6

Mazefsky, C. A., & Oswald, D. P. (2007). Emotion perception in Asperger's syndrome and high-functioning autism: The importance of diagnostic criteria and cue intensity. *Journal of Autism and Developmental Disorders*, *37*(6), 1086–1095. https://doi.org/10.1007/s10803-006-0251-6

Mendoza, E., Valencia, N., Muñoz, J., & Trujillo, H. (1996). Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS). *Journal of Voice*, *10*(1), 59–66. https://doi.org/10.1016/S0892-1997(96)80019-1

Meyer-Lindenberg, A., Domes, G., Kirsch, P., & Heinrichs, M. (2011). Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nature Reviews Neuroscience*, *12*(9), 524–538. https://doi.org/10.1038/nrn3044

Modahl, C., Green, L. A., Fein, D., Morris, M., Waterhouse, L., Feinstein, C., & Levin, H. (1998). Plasma oxytocin levels in autistic children. *Biological Psychiatry*, *43*(4), 270–277. https://doi.org/10.1016/S0006-3223(97)00439-3

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). *Physical Therapy*, *89*(9), 873–880. https://doi.org/10.1371/journal.pmed.1000097

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., & Sams, M. (2006). Perceiving identical sounds as speech or non-

speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage*, *30*(2), 563–569. https://doi.org/10.1016/j.neuroimage.2005.10.002

Mueller, J. L., Friederici, A. D., & Mannel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences*, *109*(39), 15953–15958. https://doi.org/10.1073/pnas.1204319109

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, *41*(1), 233–243. https://doi.org/10.1016/S0167-6393(02)00106-1

Nordahl, C. W., Scholz, R., Yang, X., Buonocore, M. H., Simon, T., Rogers, S., & Amaral, D. G. (2012). Increased rate of amygdala growth in children aged 2 to 4 years with autism spectrum disorders: a longitudinal study. *Archives of General Psychiatry*, *69*(1), 53–61. https://doi.org/10.1001/archgenpsychiatry.2011.145

O'Nions, E., Sebastian, C. L., McCrory, E., Chantiluke, K., Happé, F., & Viding, E. (2014). Neural bases of Theory of Mind in children with autism spectrum disorders and children with conduct problems and callous-unemotional traits. *Developmental Science*, *17*(5), 786–796. https://doi.org/10.1111/desc.12167

Olson, I. R., McCoy, D., Klobusicky, E., & Ross, L. A. (2013). Social cognition and the anterior temporal lobes: A review and theoretical framework. *Social Cognitive and Affective Neuroscience*, *8*(2), 123–133. https://doi.org/10.1093/scan/nss119

Ortiz-Mantilla, S., Choe, M. sun, Flax, J., Grant, P. E., & Benasich, A. A. (2010). Associations between the size of the amygdala in infancy and language abilities during the preschool years in normally developing children. *NeuroImage*, *49*(3), 2791–2799. https://doi.org/10.1016/j.neuroimage.2009.10.029

Pakarinen, S., Sokka, L., Leinikka, M., Henelius, A., Korpela, J., & Huotilainen, M. (2014). Neuroscience Letters Fast determination of MMN and P3a responses to linguistically and emotionally relevant changes in pseudoword stimuli. *Neuroscience Letters*, *577*, 28–33. https://doi.org/10.1016/j.neulet.2014.06.004

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Paul, R., Chawarska, K., Fowler, C., Cicchetti, D., & Volkmar, F. (2007). "Listen My Children and You Shall Hear": Auditory Preferences in Toddlers With Autism Spectrum Disorders. *Journal of Speech Language and Hearing Research*, *50*(5), 1350. https://doi.org/10.1044/1092-4388(2007/094)

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice Modulation: A Window into the Origins of Human Vocal Control? *Trends in Cognitive Sciences*, *20*(4), 304–318. https://doi.org/10.1016/j.tics.2016.01.002

Polich, J. (2003). Theoretical Overview of P3a and P3b. *Detection of Change*, 83–98. https://doi.org/10.1007/978-1-4615-0294-4_5

Quattrocki, E., & Friston, K. (2014). Autism, oxytocin and interoception. *Neuroscience and Biobehavioral Reviews*, *47*, 410–430.

https://doi.org/10.1016/j.neubiorev.2014.09.012

Rosenblau, G., Kliemann, D., Dziobek, I., & Heekeren, H. R. (2016). Emotional prosody processing in Autism Spectrum Disorder. *Social Cognitive and Affective Neuroscience*, (October), nsw118. https://doi.org/10.1093/scan/nsw118

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., … Cohen, D. (2013). Motherese in Interaction: At the Cross-Road of Emotion and Cognition? (A Systematic Review). *PLoS ONE*, *8*(10), 1–17. https://doi.org/10.1371/journal.pone.0078103

Sakkalou, E., & Gattis, M. (2012). Infants infer intentions from prosody. *Cognitive Development*, *27*(1), 1–16. https://doi.org/10.1016/j.cogdev.2011.08.003

Samson, F., Hyde, K. L., Bertone, A., Soulières, I., Mendrek, A., Ahad, P., … Zeffiro, T. A. (2011). Atypical processing of auditory temporal complexity in autistics. *Neuropsychologia*, *49*(3), 546–555. https://doi.org/10.1016/j.neuropsychologia.2010.12.033

Schelinski, S., Borowiak, K., & von Kriegstein, K. (2016). Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Social Cognitive and Affective Neuroscience*, (February), 1–11. https://doi.org/10.1093/scan/nsw089

Schelinski, S., Roswandowitz, C., & von Kriegstein, K. (2017). Voice identity processing in autism spectrum disorder. *Autism Research*, *10*(1), 155–168. https://doi.org/10.1002/aur.1639

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Schumann, C. M., Barnes, C. C., Lord, C., & Courchesne, E. (2009a). Amygdala Enlargement in Toddlers with Autism Related to Severity of Social and Communication Impairments. *Biological Psychiatry*, *66*(10), 942–949. https://doi.org/10.1016/j.biopsych.2009.07.007

Schumann, C. M., Barnes, C. C., Lord, C., & Courchesne, E. (2009b). Amygdala Enlargement in Toddlers with Autism Related to Severity of Social and Communication Impairments. *Biological Psychiatry*, *66*(10), 942–949. https://doi.org/10.1016/j.biopsych.2009.07.007

Searle, J. R. (1984). *Minds, brains and science*, pp., 43. Harvard University Press.

Seltzer, L. J., Ziegler, T. E., & Pollak, S. D. (2010). Social vocalizations can release oxytocin in humans. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1694), 2661–2666. https://doi.org/10.1098/rspb.2010.0567

Sharda, M., Midha, R., Malik, S., Mukerji, S., & Singh, N. C. (2015). Fronto-Temporal connectivity is preserved during sung but not spoken word listening, across the autism spectrum. *Autism Research*, *8*(2), 174–186. https://doi.org/10.1002/aur.1437

Sharma, A., Glick, H., Deeves, E., & Duncan, E. (2015). The P1 biomarker for

assessing cortical maturation in pediatric hearing loss: a review. *Otorinolaringologia*, *65*(December), 103–114. Retrieved from file:///C:/Users/schierir/Downloads/P1 review paper.pdf

Shen, M. D., Li, D. D., Keown, C. L., Lee, A., Johnson, R. T., Angkustsiri, K., … Nordahl, C. W. (2016). Functional Connectivity of the Amygdala Is Disrupted in Preschool-Aged Children With Autism Spectrum Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *55*(9), 817–824. https://doi.org/10.1016/j.jaac.2016.05.020

Shultz, S., Vouloumanos, A., & Pelphrey, K. (2012). The superior temporal sulcus differentiates communicative and noncommunicative auditory signals. *Journal of Cognitive Neuroscience*, *24*, 1224–1232. https://doi.org/10.1162/jocn_a_00208

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, *17*(5), 323–332. https://doi.org/10.1038/nrn.2016.23

Speer, S. R., & Ito, K. (2009). Prosody in First Language Acquisition – Acquiring Intonation as a Tool to Organize Information in Conversation, *1*, 90–110.

Stewart, M. E., McAdam, C., Ota, M., Peppé, S., & Cleland, J. (2013). Emotional recognition in autism spectrum conditions from voices and faces. *Autism*, *17*(1), 6–14. https://doi.org/10.1177/1362361311424572

Swanepoel, R., Oosthuizen, D. J. J., & Hanekom, J. J. (2012). The relative importance of spectral cues for vowel recognition in severe noise, *132*(4).

Tesink, C. M. J. Y., Buitelaar, J. K., Petersson, K. M., Van Der Gaag, R. J., Kan, C. C., Tendolkar, I., & Hagoort, P. (2009). Neural correlates of pragmatic language comprehension in autism spectrum disorders. *Brain*, *132*(7), 1941–1952. https://doi.org/10.1093/brain/awp103

Tremblay, P., Baroni, M., & Hasson, U. (2013). Processing of speech and non-speech sounds in the supratemporal plane: Auditory input preference does not predict sensitivity to statistical structure. *NeuroImage*, *66*, 318–332. https://doi.org/10.1016/j.neuroimage.2012.10.055

Tsao, L. (2008). Social , Language , and Play Behaviors of Children with Autism, *14*, 40–51.

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: A Dissociation Between Familiar and Unfamiliar Voices. *Cortex*, *24*(2), 195–209. https://doi.org/10.1016/S0010-9452(88)80029-7

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of Sounds in the Auditory Stream: Event-Related fMRI Evidence for Differential Activation to Speech and Nonspeech. *Journal of Cognitive Neuroscience*, *13*(7), 994–1005. https://doi.org/10.1162/089892901753165890

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x

Wang, J. E., & Tsao, F. M. (2015). Emotional prosody perception and its association with pragmatic language in school-aged children with high-function autism. *Research in Developmental Disabilities*, *37*(1), 162–170. https://doi.org/10.1016/j.ridd.2014.11.013

Wartenburger, I., Steinbrink, J., Telkemeyer, S., Friedrich, M., Friederici, A. D., & Obrig, H. (2007). The processing of prosody: Evidence of interhemispheric specialization at the age of four. *NeuroImage*, *34*(1), 416–425. https://doi.org/10.1016/j.neuroimage.2006.09.009

Whitehouse, A. J. O., & Bishop, D. V. M. (2008). Do children with autism "switch off" to speech sounds? An investigation using event-related potentials. *Developmental Science*, *11*(4), 516–524. https://doi.org/10.1111/j.1467-7687.2008.00697.x

Yatawara, C. J., Einfeld, S. L., Hickie, I. B., Davenport, T. A., & Guastella, A. J. (2016). The effect of oxytocin nasal spray on social interaction deficits observed in young children with autism: a randomized clinical crossover trial. *Molecular Psychiatry*, *21*(9), 1225–1231. https://doi.org/10.1038/mp.2015.162

Yoshimatsu, Y., Umino, A., & Dammeyer, J. (2016). Characteristics of the Understanding and Expression of Emotional Prosody among Children with Autism Spectrum Disorder. *Autism Open Access*, *6*(4), 100185. https://doi.org/10.4172/2165-7890.1000185

Yoshimura, Y., Kikuchi, M., Hiraishi, H., Hasegawa, C., Takahashi, T., Remijn, G. B., … Kojima, H. (2016). Atypical development of the central auditory system in young children with Autism spectrum disorder. *Autism Research*, *9*(11), 1216–1226. https://doi.org/10.1002/aur.1604

Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., & Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences*, *104*(15), 6430–6435. https://doi.org/10.1073/pnas.0607061104

Zalla, T., & Sperduti, M. (2013). The amygdala and the relevance detection theory of autism: an evolutionary perspective. *Frontiers in Human Neuroscience*, *7*(December). https://doi.org/10.3389/fnhum.2013.00894

Zatorre, R. J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex, 946–953.

Zhang, R., Zhang, H. F., Han, J. S., & Han, S. P. (2017). Genes Related to Oxytocin and Arginine-Vasopressin Pathways: Associations with Autism Spectrum Disorders. *Neuroscience Bulletin*, *33*(2), 238–246. https://doi.org/10.1007/s12264-017-0120-7

Zilbovicius, M., Saitovitch, A., Popa, T., Rechtman, E., Diamandis, L., Chabane, N., … Boddaert, N. (2013). Autism , social cognition and superior temporal sulcus. *Open Journal of Psychiatry*, *2013*(April), 46–55. https://doi.org/10.4236/ojpsych.2013.32A008

Appendix

Cortical area	Function
Superior Temporal Gyrus/Sulcus	- The voice-sensitive areas in the auditory cortex responsible for the recognition of vocal sounds (Belin et al., 2004) - The representation of auditory percept after a low-level acoustic analysis within the primary auditory cortex (Leitman et al., 2016; Fruhholz et al., 2016) - Activated when emotionally significant acoustic information is perceived so further emotional evaluation can be done (Schirmer & Kotz, 2006)
Inferior Frontal Gyrus	- Activated during the explicit judgment of emotional prosody, predominantly in the right hemisphere (Schirmer & Kotz, 2006) - Contributes to altered comprehension of vocal emotions when lesioned (Ethofer, 2012), important for the detection of emotional vs. neutral tone of voice (Alba-Ferrara, 2011) - Involved in forming and updating a situation model, i.e. integration of implausible or unexpected information given the current situation model and general world knowledge (Tesink et al., 2009)
Orbitofrontal Cortex	- Involved during explicit evaluative judgments of emotional prosody together with the inferior frontal gyrus (Schirmer & Kotz, 2006) - Involved in processing of emotionally salient affective stimuli such as nonverbal vocalizations (Blasi et al., 2011) - Functionally connected to the amygdala, and together with the superior temporal gyrus form the "social brain" (Zalla & Sperduti, 2013)
Amygdala	- Detects salient and personally relevant stimuli in cooperation with ventral and dorsal medial prefrontal cortex. Amygdala and medial prefrontal cortex are also engaged (Fossati, 2012) - One of the most important subcortical areas playing crucial role for vocal affect information together with temporo-medial regions, anterior insula and inferior prefrontal regions predominantly in the right hemisphere (Belin et al., 2011) - Plays a critical role in automatic emotional response behavior, which deactivation is observed during explicit task instruction (Kotz, et al., 2012)

Table 5: A brief description of relevant cortical and subcortical areas for the processing of different types of information that a human voice may convey.